1999

Synthesis, Thermodynamic Stability and Enzymic Behavior of Oligonucleotides Containing Pyrazole Nucleobase Analogs.

Andrea Sapp Saurage
Louisiana State University and Agricultural & Mechanical College

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600

SYNTHESIS, THERMODYNAMIC STABILITY AND ENZYMIC BEHAVIOR OF OLIGONUCLEOTIDES CONTAINING PYRAZOLE NUCLEOBASE ANALOGS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Chemistry

by
Andrea Sapp Saurage
B.S., North Carolina State University, 1993
B.S., North Carolina State University, 1994
December 1999

Dedication

This collection of my work over the last four years is dedicated to my parents. This is a culmination of over 23 years of education provided by my parents. They have constantly provided love, encouragement, strength and support throughout all of my endeavors. I also dedicate this to my husband, Kevin. He endured the daily aspects of the degree with continuous love and support. I would not have been able to finish without his encouragement and positive attitude. To my son Evan, I hope this will serve as an inspiration to work hard for what you want and to always finish what you start. To the rest my family, thank you for the continuous support, love, and encouragement you have provided.

Acknowledgments

First, I would like to thank Prof. Robert Hammer for his first phone call to me over 4 years ago. My life would have been very different without it. He has been not only a wealth of knowledge, but a great friend and advisor. Next I would like to thank the two postdocs, Dr. Kris Moffett and Dr. Tod Miller. They have provided many hours of discussion and guidance over the years, both in and out of the lab. Dr. Tod Miller also provided an "extra" set of hands during and after my pregnancy. Without him, I would not have been able to finish. The rest of the research group has also been very helpful in my progress. The first being Dr. Melissa Cameron who began the work in the thermodynamic stability and the synthesis of the pyrazole compounds. Hannah Andrepont Farquar has also been a valuable member, she helped with the loose ends of the DNA polymerase project. Glenna Sowa provided starting material for making nucleosides and general support for various aspects of the project. I would also like to thank the rest of my committee members, the faculty, and the staff of the Chemistry Department for their assistance over the years. Dr. Geoff Hoops, Prof. Jo Davisson, Prof. Donald Bergstrom and Natasha Paul (Ph.D. candidate) at Purdue University have provided many helpful discussions and shared unpublished research results and protocols that have allowed the DNA polymerase project to proceed. Prof. Kathleen Morden provided much needed guidance and instrumentation for our thermodynamic melting studies. Prof. Steve Soper provided space in his research lab for the DNA polymerase project that required the use of radioactive materials. Prof. Sue Bartlett provided access to the

STORM scanner and software to analyze our data for the DNA polymerase project. Without these people and many others in the department, the research presented herein would have been much more difficult to complete.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Abbreviations

Abbreviation	Name
Me	Methyl
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
Con	Control
dTTP	Deoxythymidine triphosphate
dATP	Deoxyadenosine triphosphate
α	Mole fraction of single stranded DNA in the duplex state
A	Deoxyadenosine
AcOH	Acetic acid
bp	Base pair
bz	Benzoyl
C	Deoxycytidine
Conc.	Concentration
CPG	Control pore glass
CSO	(1S)-(+)-(10-camphorsulfonyl)-oxaziridine
DCM	Dichloromethane
dNTP	Deoxynucleotide triphosphate
ddNTP	Dideoxynucleotide triphosphate
DFT	2,4-Difluorotoluene

Abbreviation	Name
DIEA	Diisopropyylethylamine
DMAP	4,4-Dimethylaminopyridine
DMF	Dimethylformamide
DMT	Dimethoxytrityl
DMT-Cl	Dimethoxytrityl chloride
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphospate
dsDNA	Double stranded DNA
EDTA	Ethylenediaminetetraacetic acid
Et_3N	Triethylamine
EtOAc	Ethyl acetate
EtOH	Ethanol
FABMS	Fast atom bombardment mass spectrometry
G	Deoxygaunosine
HR FAB	High resolution fast atom bombardment mass spectrometry
MALDI-MS	Matrix assisted laser desorption ionization mass spectrometry
MCPBA	*m*-Chloroperoxybenzoic acid
MeOH	Methanol
m.p.	Melting point
NMR	Nuclear Magnetic Resonance

Abbreviation	Name
PCR	Polymerase Chain Reaction
R_f	Ratio of the distance traveled by a compound relative to the solvent front
RNA	Ribonucleic acid
ssDNA	Single stranded DNA
T	Deoxythimidine
TAE	Tris-acetic acid EDTA buffer
TBE	Tris-boric acid-EDTA buffer
TEAA	Triethyammonium acetate
TEAB	Triethyammonium bicarbonate
THF	Tetrahydrofuran
TLC	Thin layer chromatography
T_m	Melting temperature
Tol	*p*-toluoyl
UV	Ultraviolet

Abstract

Non-hydrogen bonding nucleosides with iodo, nitro, propynyl and thiazolyl substituted at the 4-position of pyrazole were prepared. These nucleosides were converted to their corresponding nucleoside phosphoramidites and incorporated into a series of complementary oligonucleotides in order to determine the effect that varying size, charge distribution and polarizability has on duplex stability and structure. The self-complementary Dickerson dodecamer sequence 5'-CGCXAATTYGCG-3', as well as the non-self complementary sequence 5'-CAAAATGGTGGCCAAGT-3' previously investigated by Brown, were used to determine the duplex stabilization and thermodynamic consequences of placing 5-nitroindole and 4-substituted pyrazole nucleosides across from each natural base. In the pyrazole series, the largest and smallest duplex destabilization in all cases studied was found with cytosine and adenosine, respectively, while the 4-thiazolyl substitution was determined to form the most stable duplex regardless of the complementary base. We investigated the directing ability of the pyrazole analogs for incorporation of dNTP's. The DNA polymerases chosen were *Taq* and Pfu exo-, which lack 3'-5' proofreading exonuclease activity, and Vent, Deep Vent, Pfu which contain exonuclease activity. The enzymes were able to incorporate natural nucleotides across from our modification as detected using the polymerase chain reaction. The Sanger method of dideoxysequencing was used to determine the natural base incorporated across from the modification. The non-proofreading

enzymes mainly incorporated deoxyadenosine across from our modification where as the proofreading enzymes removed our modification but not with total efficiency.

Automated DNA synthesis procedures might modifying the 5-nitroindole base via electrophilic aromatic substitution by replacing the nitro group or addition of iodine to the ring system. An alternative oxidation reagent, CSO, was used to prepare a series of sequences for comparison with sequences prepared using standard oxidation conditions. Results from MALDI-MS did not provide conclusive evidence for either substitution or iodination of the bicyclic ring system.

Examination of the thermodynamic results led to the preparation of thiazole-C-nucleoside, which historically is more difficult to synthesize. The thiazole ring offers a site for the formation of an N-oxide. It has been suggested this group is responsible for enzyme recognition. We prepared the thiazole and thiazole-N-oxide nucleosides for incorporation into oligonucleotides for thermodynamic and enzymic investigations.

Chapter 1
Introduction
Non-hydrogen Bonding Nucleosides

1.1 Background

The design and synthesis of non-natural base analogs was originally sought after in hopes of a universal base. A universal base is defined as one that would not appreciably destabilize the duplex and would pair equally wee with both purines and pyrimidines. Also, enzymes would recognize and incorporate equal amounts of the natural nucleotides (dNTPs) across from the universal base. Incorporation of all bases across from a universal base would allow random mutagensis at directed sites.

3-Nitropyrrole was one of the first bases to show almost universal thermodynamic stability when paired with any of the natural bases.[1] However, polymerase reaction with 3-nitropyrrole showed a large preference for incorporation of purines over pyrimidines.[2] Thus another application of this type of nucleoside analogs is for directed conversion of one base to another.

Often mismatches have to be created when attempting to convert one base pair to another (e. g. GC bp → AT bp). A G-T "wobble" mismatch in the primer results in reduced polymerase extension efficiency (poor read) and thus low efficiency of conversion. As shown with the use of a non-natural base could improve this by eliminating the poor read step as the analog could pair well (Watson-Crick bonding pattern) with the natural base, unlike the wobble mismatch. The resulting "good read" would improve the overall conversion of the sequence.

One application of polymerase-based conversion with nucleobase analogs is creation of restriction sites in DNA for enhancing detection of low level mutations

in a large background of wild-type (non-mutated) DNA.[3,4] The introduction of non-natural base into a PCR primer will allow for conversion of the sequence to generate a restriction site in wild type DNA while the low-level mutant DNA will not posess a restriction site. Following this PCR step (repeating many times if necessary), a restriction enzyme will digest the DNA removing the converted wild-type DNA and leaving the mutant DNA uncut.[3,4]

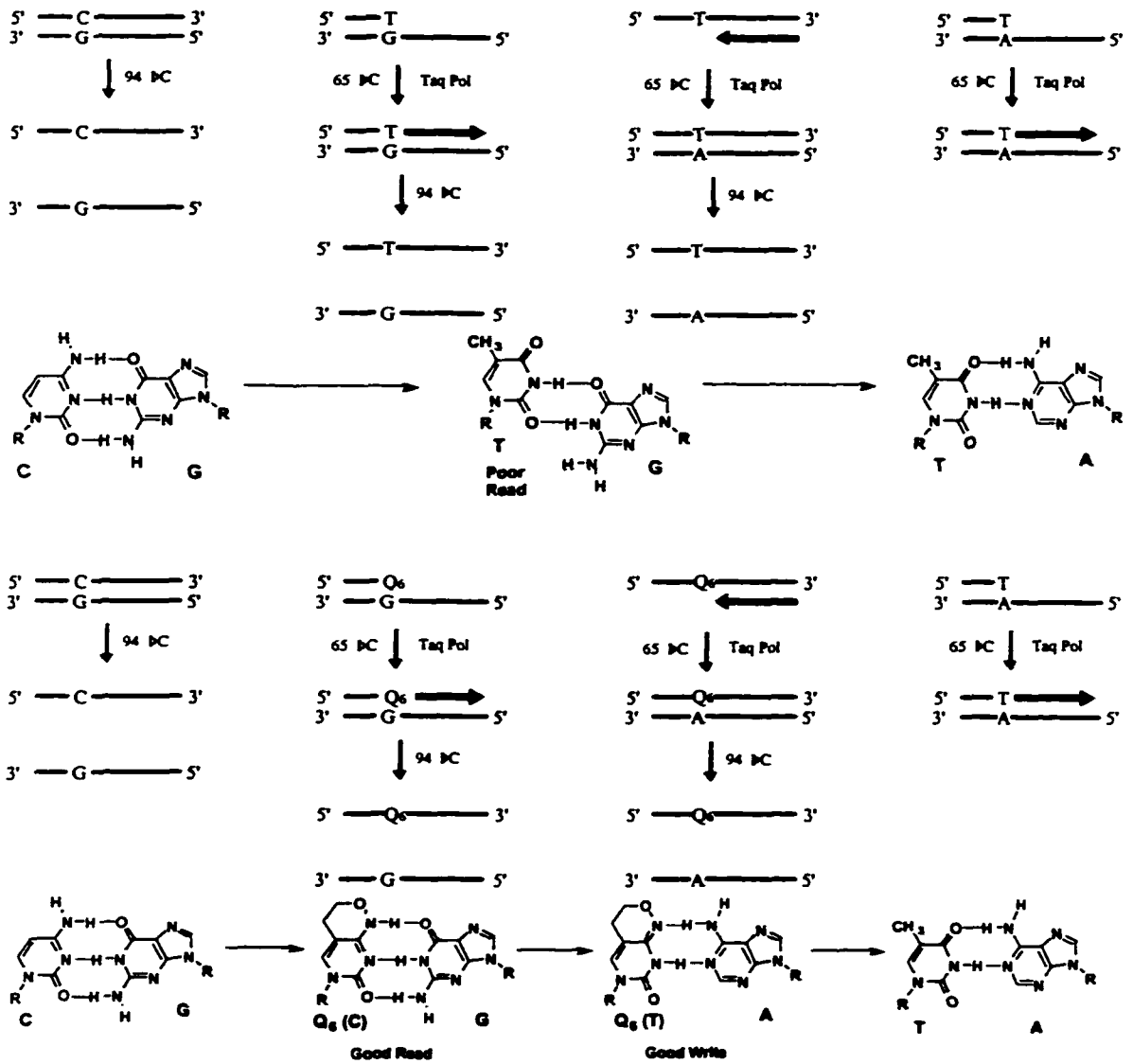


Figure 1.1 Conversion of C–G to T-A base pair

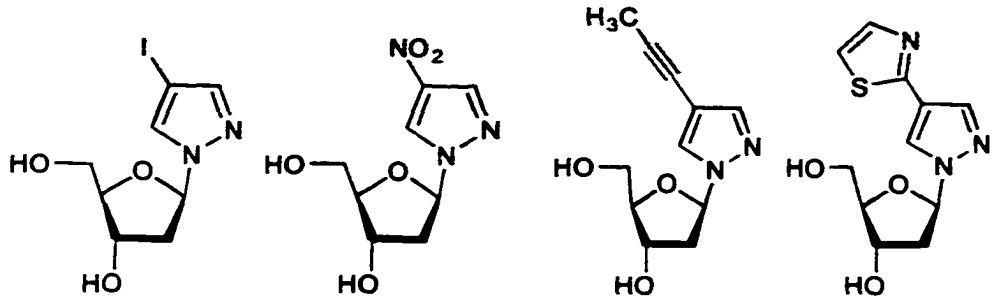
1.2 Design

The introduction of non-natural nucleosides has been under intense investigation as a way to understand the extremely efficient design of DNA and its properties. We and others are interested in dissecting the roles hydrogen bonding and "base stacking" play in controlling duplex stability, structure, and DNA enzymology. One approach to studying these factors is to remove the hydrogen bonding of the bases altogether. A first foray into this approach was replacing the heteroaromatic bases with either a hydrogen atom or a phenyl ring,[5] which found that both substitutions were similarly destabilizing to the duplex. The pioneering work of Bergstrom and coworkers used an intercalation model to develop a 3-nitropyrrole analog,[6,7] which had much great stability in a duplex relatively to the Millican phenyl or abasic site analogs.[5]

Another approach is to mimic the size and charge of the natural nucleosides as Kool and coworkers have suggested. They designed 2,4-difluorotoluene as a shape mimic of thymidine.[8,9] The methyl group is in same position and the fluoro groups are similar to the oxygens on thymidine both in size and electronegativity. Kool also designed 4,6-dimethyindole as a hydrophobic adenine analog.[8,9](Figure 1.5) The methy groups are in the proximity of the oxygens in T to offer similar shape without specific hydrogen bonding characteristics.

We have developed a series non-hydrogen bonding analogs (Figure 1.2) to investigate the properties of DNA when hydrogen bonding is removed while varying the charge and size characteristics of substituents on the aromatic core. We designed a series of pyrazole analogs with substituents in the 4-position with

varying charge and size, iodo, propynyl, nitro and thiazolyl (**2a, 2b, 2c, and 2e**). The differing characteristics of the substituents allows for investigation of charge distribution and size variation on the stability of DNA duplexes as well as the enzymic acceptance of the analogs. The size of the small 5-membered should fit into the active site of enzyme and not offer steric hindrance when incorporated across from purines.



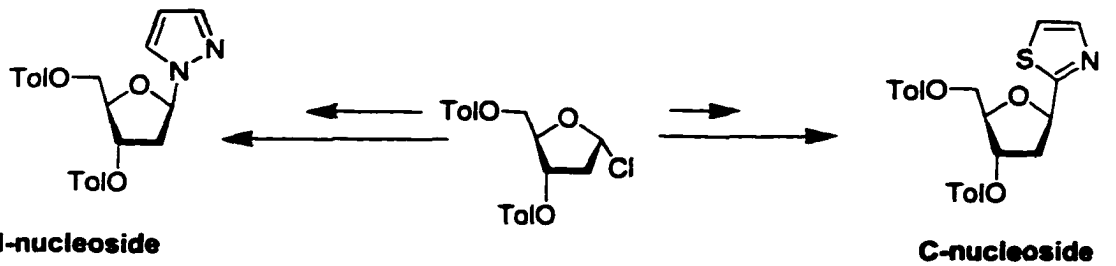
PzI **2a** **PzN** **2b** **PzP** **2c** **PzT** **2e**

Figure 1.2 Series of pyrazole analogs

1.3 Synthesis

Once the series of pyrazole analogs was chosen, the synthesis of these non-natural analogs had to be designed. There are two traditional approaches for the synthesis of nucleosides. The first and more popular approach uses the corresponding base analog reacting with a sugar resulting in the nucleoside. This type of reaction is called glycoslyation. The Hilbert-Johnson method is a classic glycosylation method used in the synthesis of N1-glycosyl pyrimidines.[10] The second approach requires the construction of the base analog system from an N or C-glycoslated precursor. There are 2 different types of nucleosides, C or N. We are interested in the synthesis N-nucleosides. Meaning, the C1 position of the sugar is attached to the bases through a nitrogen in the ring of the base (Figure 1.3). The

other type is a C-nucleoside. These have the sugar attached to a carbon in the ring of the base. C-nucleosides are much more difficult to synthesize. This difficulty arises from the unstable precursors, like glycals[11] or a low yielding multistep reactions which builds the aryl unit from a cyano or some other C-glycoside (Figure 1.3). We are interested in attaching the N1 of the commercially available pyrazole ring to the sugar precursor. The substituents are either already present on the ring when purchased (iodo and nitro) or they are synthesized (propynyl and thiazolyl) once the pyrazole ring is attached to the sugar using palladium mediated chemistry.



N-nucleoside **C-nucleoside**

Figure 1.3 N-nucleoside and C-nucleoside synthesis

1.4 Thermodynamic Stability

The arrangement of the aromatic bases in DNA orthogonally, along the center axis of the duplex allows pyrimidine-purine hydrogen bonding between the two strands (edge-edge interactions) as well as stacking of bases with the bases above and below it in the duplex (π-π interactions). After removal of hydrogen bonding, a heterobase analog utilizes mainly π-stacking along the axis of the duplex, though dipole-dipole interactions across the duplex with its putative "pairing" partner may also influence structure and stability of the duplex.

One factor that affects stacking ability is the charge distribution in the π-aromatic system of the analog. For example, 3-nitropyrrole (Figure 1.5) is much

less destabilizing than a phenyl ring.[12] Bergstrom showed that an unsubstituted pyrazole ring which has some charge distribution is equivalent to an abasic site.[13] The addition of a nitro group to the 4-position of the pyrazole ring dramatically increases the stability of the helix by 10-20 °C.[13]

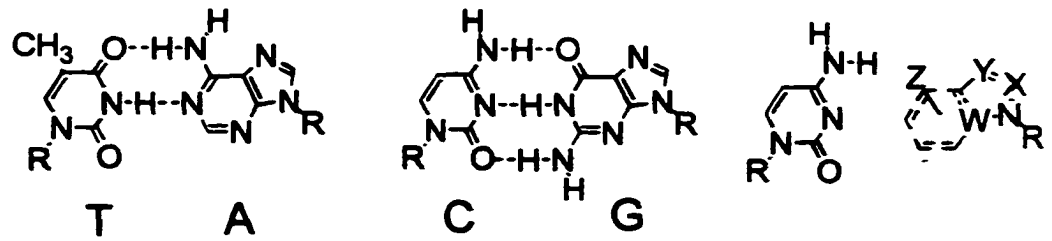
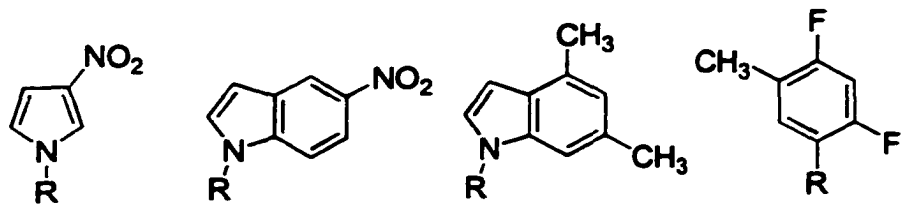


Figure 1.4. Comparison of hydrogen bonding base pair with a non-hydrogen bonding analog.

A second factor affecting stacking ability is size. This is demonstrated by the increased stability of duplexes with 5-nitroindole (Figure 1.5) over 3-nitropyrrole by approximately 20 °C.[13] In a similar vein, Froehler has shown propynyl[14] and thiazolyl[15] substituents, on hydrogen bonding bases, stacking in the major groove can increase the duplex T_m by 1-2°C per substitution. (Figure 1.6)



3-Nitropyrrole 5-Nitroindole 4,6- Dimethylindole 2,4-Difluorotoluene

Figure 1.5 Structures of analogs studied by Bergstrom,[6] Brown,[12] and Kool[9]

The size and charge distribution of the non-hydrogen-bonding analog also has an effect on the preferred pairing partner when investigating duplex stability. For example, Bergstrom and coworkers have found 3-nitropyrrole base analog to be less destabilizing to the duplex when "paired" with deoxyadenosine (A) and

deoxygaunosine (G).[6] Whereas, the 5-nitroindole nucleoside base analog is less destabilizing when paired with deoxycytidine (C).[12] Kool has investigated several analogs (Figure 1.5) such as 4,6-dimethylindole and 2,4-difluorotoluene, while diflourotoluene is a deoxythimidine (T) analog it is least stable across from A and in fact is most stable across from itself in the duplex.[9]

To determine the stacking ability of a variety of substituents without hydrogen bonding effects, we have prepared a series of 4-substituted pyrazole nucleoside analogs and studied their effects on DNA duplex stability in a variety of contexts.

1.5 Enzymic Activity

1.5.1 DNA Polymerases

DNA polymerases are responsible for the replication of DNA. Kornberg isolated the first DNA polymerase in 1955 from an extract of *E. coli*. This polymerase, now called DNA polymerase I, has two main segments with three main functions. The N- terminal small fragment contains 5'→3' exonuclease activity. Exonuclease activity is the ability of a polymerase to remove an imperfect match. The large fragment, better known as the Klenow fragment, has polymerase activity as well as 3'→5' exonuclease activity. These three functions are crucial to correct replication of DNA.

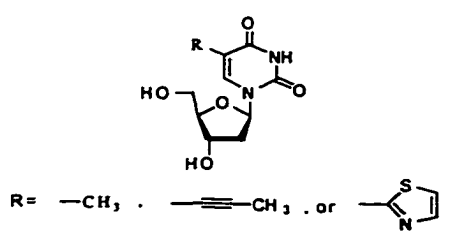


Figure 1.6 Structures of thymidine, 5-(1-propynyl)-2'deoxyuridine and 5-(2-thiazoyl)-2'-deoxyuridine

Replication of DNA, also called extension of a primer, involves several components. First, the enzyme must posses polymerase activity. The sequence to be copied, called the template, must be present. Next, a primer (much shorter in length) matching a portion of the template sequence must anneal to the template making a double strand of DNA. For the polymerase to extend off the 3'-end of the primer, deoxynucleotide triphophates must be availble to the polymerase and an optimal temperature reached for extension to begin. Once these main components are present, replication of the template is possible.

1.5.2 Polymerase Chain Reaction (PCR)

The introduction of the polymerase chain reaction (PCR) revolutionized DNA extension technology.[16,17](Figure 1.8) This tool allowed for rapid amplification of fmol quantities of DNA. Whereas the previous extension technology (Figure 1.7) only provided linear amplification. The concept is similar to that described in Section 1.5.1 but the cycle is repeated many times varying the temperature to allow for annealing of the primer, extension and denaturing of the double stranded product. Also the product is increased exponentially due to the presence of a double stranded template and a primer for each strand of the double-stranded template. The extension explained earlier only required a single stranded template and one primer. The desired double stranded sequence to be amplified is called the template. The template is normally present in fmol quantities. At least some of the template sequence must be known so as to offer a place for annealing of the primer so that amplification can occur. The primers serve as the starting place for the DNA polymerase to start copying the template. A primer must be designed

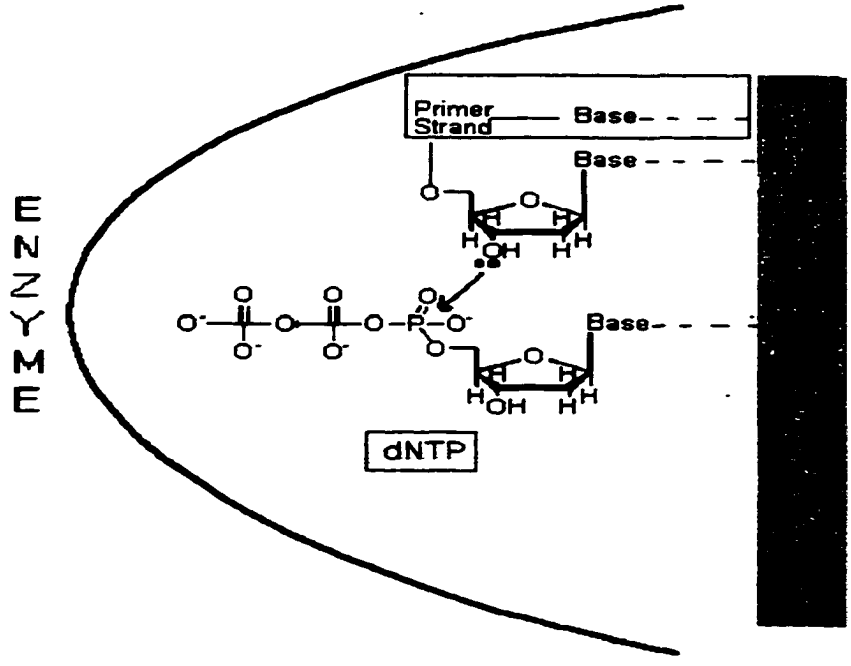


Figure 1.7 Polymerase extension

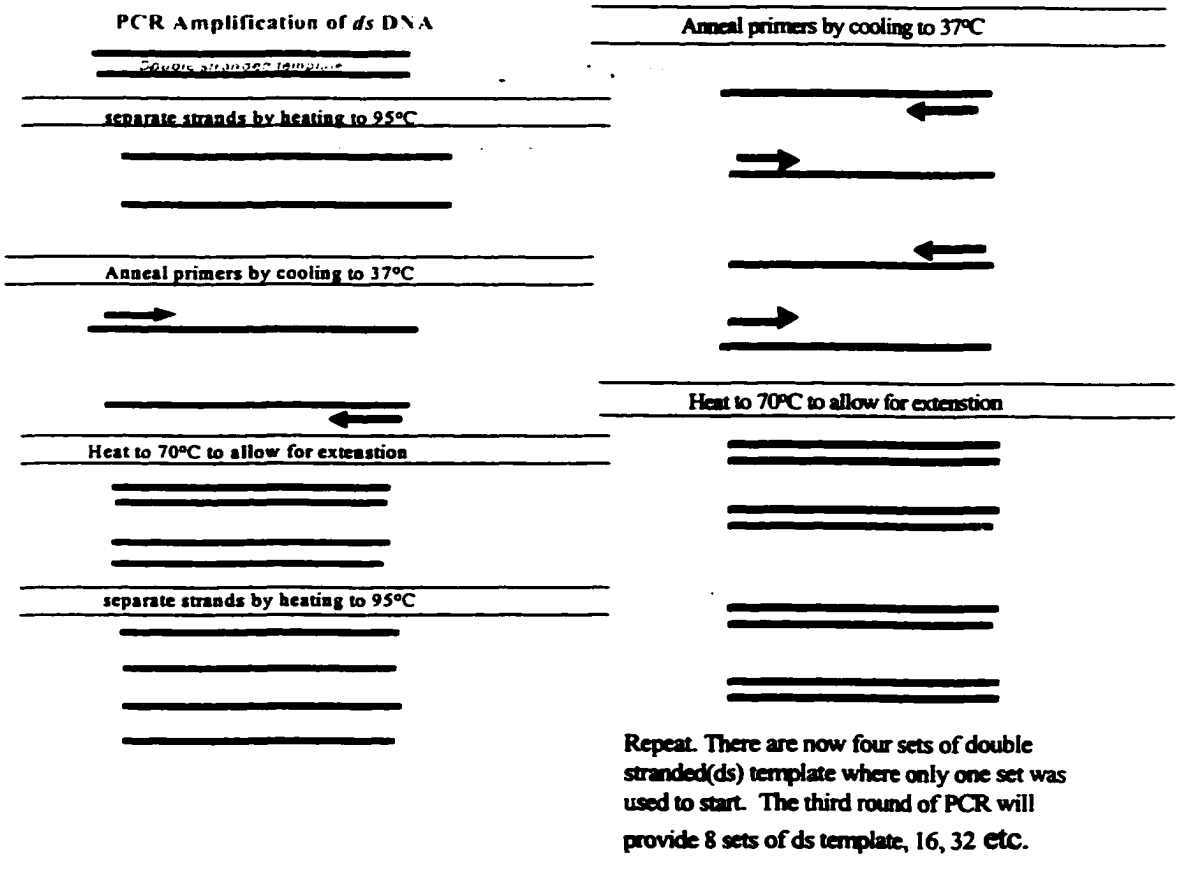


Figure 1.8 PCR amplification cycle

for each strand of the template. Once the reaction starts, the primers are extended in the 3' →5' direction to the end of both strands of the template. The polymerase inserts the complementary dNTP that creates a natural base pair with the template strand and then continues on to the next base. At the completion of this cycle, a copy of the original template is created. The double stranded products are separated by heating to 95 °C and now a place is available for new primers to be positioned for extension. To position the primers on the template an annealing step is performed (normally 50-60 °C) followed by raising the temperature to 70-75 °C (depending on polymerase) for optimal extension temperature. On the second cycle, copies of old and new pieces of template are made resulting in a total of four templates. The third cycle copies the four strands resulting in a total of eight and so the template is amplified exponentialyl.

The Klenow fragment was not particularly efficient for PCR as the high heat required in thermal cycling kills the polymerase activity of the enzyme. Therefore the discovery of thermostable polymerase was necessary for the success of PCR. The thermostable polymerase under the most intense investigation was *Thermus aquaticus* (*Taq*).[18-21] Other thermostable polymerases that are commonly used are *Thermotoga maritima* (*Tma*)[22], *Pyrococcus furiosus* (*Pfu*)[23-25], *Thermus thermophilus* (*Tth*) [26,27], *Pyrococcus species* GB-D [28] (Deep Vent from New England Biolabs) and *Thermococcus litoralis* (*Tli*) [29] (Vent from New England Biolabs). These polymerases as well as many others provide different tools for investigating modified primers, modified triphosphates or modified templates.

1.5.3 Sanger Method for Sequencing DNA

The Sanger method is very similar to PCR in the fact that a polymerase, a primer, a template and dNTP's are used. (Figure 1.9) This technique only provides linear amplification of the desired product. What makes this method different is there are dideoxy nucleotide triphosphates (ddNTP's) present. ddNTP's lack the 3´ hydroxyl group removing the active group for extension. This creates truncated pieces that have been extended somewhat but once the ddNTP is incorporated no extension is possible. For the convenience of determining the ddNTP that terminates the strand the ddNTP's are radiolabeled with ^{33}P for the purpose of imaging the truncated sequences. (Figure 1.9) Each of the four natural bases are used as ddNTPs but in separate reaction so as to be able to determine the exact sequence. Once the extension reaction is complete, the truncated pieces are loaded onto a polyacryamide gel to be separated using gel electrophoresesis based on differences in size. The shorter pieces move at a faster rate through the gel as the longer pieces move at a slower rate. Once the electrophoresis is complete, a ladder of the different sized pieces is present. Each separate reaction with an individual ddNTP is loaded in individual lanes and separated at the same time. This creates a ladder of the different sized pieces and allows for determination of the sequence.(Figure 1.9) An alternative method for sequencing is the Maximum-Gilbert method. This required chemical degradation with regents specific for one or two of the natural bases followed removal of the base creating an abasic site along the backbone of the DNA. Once the abasic cite is present cleavage of the backbone is much easier resulting in truncated sequences as well.

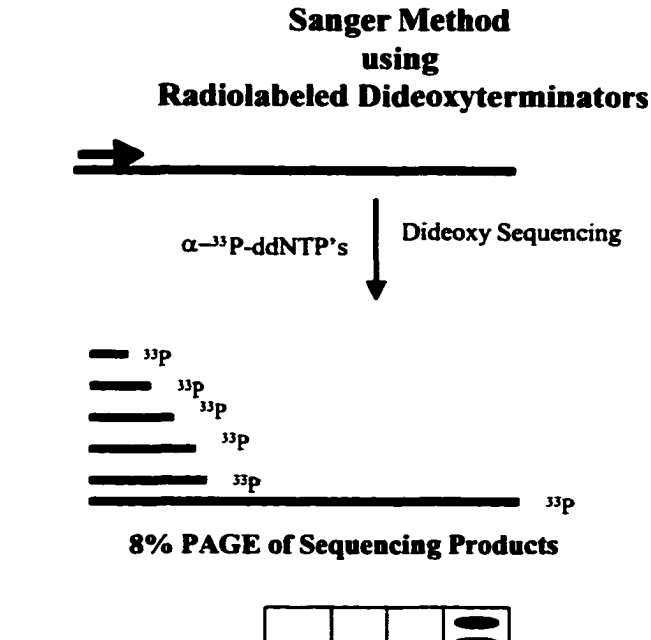


Figure 1.9 Sanger method for dideoxy sequencing

1.5.4 Exonuclease Activity

There are two directions for exonuclease activity to occur 3′→5′ and 5′→3′. 5′→3′ Exonuclease is useful in removing misprimed sequences and to repair damaged nucleotides.[30] The 3′→5′ activity can remove unnatural modification by cutting up the DNA until the unnatural analog is removed. With 3'→5' exonuclease activity, also known as proofreading, the enzyme can remove an incorrectly inserted nucleotide at the primer terminus or 3'-end before continuing with polymerization.(Figure 1.10) The mismatched nucleotide is not involved in hydrogen bonding with the template strand. The ability to proofread drastically

increases the accuracy of the DNA replication by allowing the polymerase to correct any mistakes before continuing.

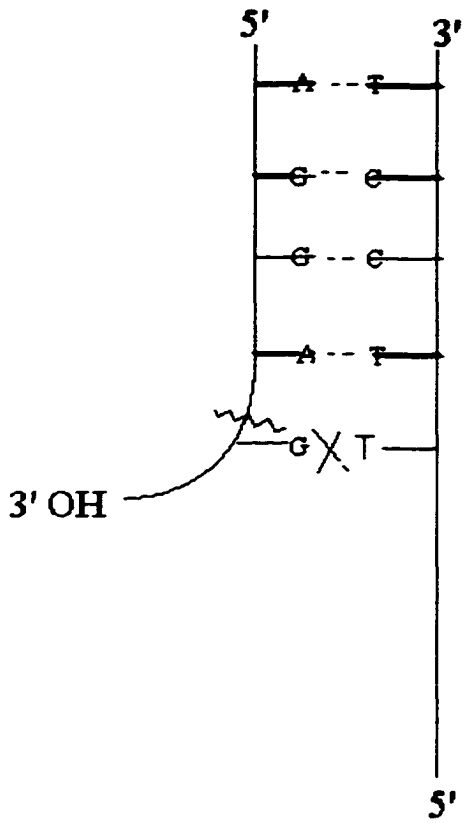


Figure 1.10 3'→5' Exonuclease activity –proofreading

5'→ 3' exonuclease activity allows the polymerase to excise any misprimed sequences and continue with polymerization. (Figure 1.11) Without this feature, the polymerase would discontinue polymerization of the template when misprimed sequences were found (most often in nature with RNA primers). Without this activity, DNA replication is not as accurate and complete replication of the template is hindered.

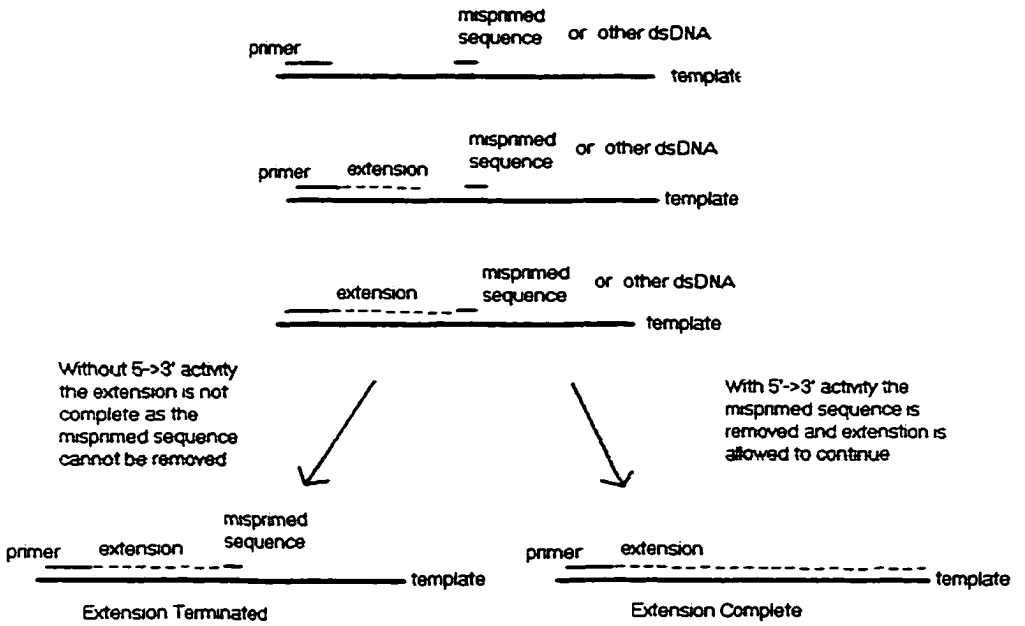


Figure 1.11 5'→3' Exonuclease activity -misprimed sequences

1.5.5 Polymerase Reactions with Non-Natural Nucleosides

The behavior of enzymes with non-natural nucleosides has drawn much attention and research. Bergstrom reported one of the first successful PCR reactions using a modified primer containing 3 nitropyrrole.[6,7] Other work by Bergstrom utilizing azole nucleobase analogs provided the model for our experiments.[2] Bergstrom's research studied 3-nitropyrrole, inosine, 3-pyrrolecarboxamide, 4-pyrazole-carboxamide and 4- imidazolecarboxamide, along with an abasic site with PCR utilizing *Taq* DNA polymerase. (Figure 1.12) These results showed a 3:1 A:T incorporation in PCR for 3-nitropyrrole using the Sanger sequencing method with ^{33}P radiolabeled dideoxyterminators. Similar research by Kool and coworkers uses 2,4-diflorotoulene (DFT) as a T mimic. Kool has successful incorporation of DFT triphosphates across from A and incorporation of dATP across from DFT analogs in

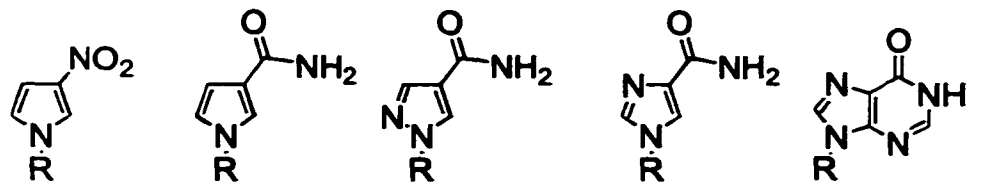
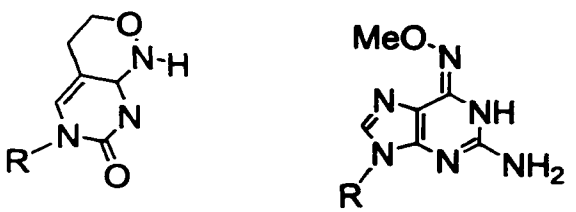


Figure 1.12 Structures of 3-nitropyrrole, 3-pyrrole-carboxamide, 4-pyrazole-carboxamide, 4-imidazole-carboxamide, and inosine (left to right)

primers.[9](Figure 1.5) Brown and coworkers have used 6H, 8H-3,4-dihydropyrimido[4,5-c][1,2]oxazin-7-one, a pyrimidine analog, and N^6-methoxy-2,6-diaminopurine, a purine analog, P and K respectively.[31](Figure 1.13) The results with these modification incorporated into a primer using *Taq* polymerase show these analog behave as designed a purine analog, K and a pyrimidine analog, P. The research of Bergstrom[2] and Brown[31] has shown the ability of *Taq* polymerase to accept a non-natural nucleobase in the DNA primer as well as non-natural triphosphates.



P K

Figure 1.13 Structures of 6H, 8H-3,4-dihydropyrimido[4,5-c][1,2]oxazin-7-one (P) and N^6-methoxy-2,6-diaminopurine (K)

Our research will investigate the acceptance of the pyrazole analogs with *Taq* DNA polymerase as well as other thermostable polymerases with a variety of inherent properties using PCR. We will also investigate how the pyrazole analogs are read by the enzymes using sequencing methods to determine the natural base incorporated across from the analogs.

1.6 References

1) Zhang, P.; Johnson, W. T.; Klewer, D.; Paul, N.; Hoops, G.; Davisson, V. J.; Bergstrom, D. E. *Nucleic Acids Res.* **1998**, *26*, 2208-2215.

2) Hoops, G. C.; Zhang, P.; Johnson, W. T.; Paul, N.; Bergstrom, D. E.; Davisson, V. J. *Nucleic Acids Res.* **1997**, *25*, 4866-4871.

3) Day, J. P.; Hammer, R. P.; Bergstrom, D.; Barany, F. *Nucleic Acids Res* **1999**, *27*, 1819-1827.

4) Day, J. P.; Bergstrom, D.; Hammer, R. P.; Barany, F. *Nucleic Acids Res* **1999**, *27*, 1810-1818.

5) Millican, T. A.; Mock, G. A.; Chauncey, M. A.; Patel, T. P.; Eaton, M. A. W.; Gunning, J.; Cutbush, S. D.; Niedle, S.; Mann, J. *Nucleic Acids Res.* **1984**, *12*, 7435-7453.

6) Nichols, R.; Andrews, P. C.; Zhang, P.; Bergstrom, D. E. *Nature* **1994**, *369*, 492-493.

7) Bergstrom, D. E.; Zhang, P.; Toma, P. H.; Andrews, P. C.; Nichols, R. *J. Am. Chem. Soc.* **1995**, *117*, 1201-1209.

8) Schweitzer, B. A.; Kool, E. T. *J. Org. Chem.* **1994**, *59*, 7238-7242.

9) Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863-1872.

10) Pliml, J.; Pyrstas, M.; . *Adv. Heterocyc. Chem* **1967**, *8*, 115.

11) Cameron, M. A.; Cush, S. B.; Hammer, R. P. *J. Org. Chem.* **1997**, *62*, 9065-9069.

12) Loakes, D. *Nucleic Acids Res.* **1994**, *22*, 4039-4043.

13) Bergstrom, D. E.; Zhang, P.; Johnson, W. T. *Nucleic Acids Res.* **1997**, *25*, 1935-1942.

14) Froehler, B. C.; Wadwani, S.; Terhorst, T. J.; Gerrard, S. R. *Tetrahedron Lett.* **1992**, *33*, 5307-5310.

15) Guiterrez, A. J.; Terhorst, T. J.; Matteucci, M. D.; Froehler, B. C. *J. Am. Chem. Soc.* **1994**, *116*, 5540-5544.

16) Mullis, K. B.; Faloona, F. *Meth. Enzymol.* **1987**, *155*, 335-350.

17) Saiki, R. K.; Gefland, D. H.; Stoffel, S.; Scharf, S.; Higuchi, R.; Horn, G. T.; Mullis, K. B.; Erlich, H. A. *Science* **1988**, *239*, 487-491.

18) Chein, A.; Edgar, D. B.; Trela, J. M. *J. Bacteriol.* **1976**, *127*, 1550-1557.

19) Kaledin, S. S.; Sliusarenko, A. G.; Gorodetskii, S. I. *Biokhimiya* **1980**, *45*, 644-651.

20) Lawyer, F. C.; Stoffel, S.; Saiki, R. K.; Myambo, K.; Drummond, R.; Gefland, D. H. *J. Biol. Chem.* **1989**, *264*, 6427-6437.

21) Longley, M. J.; Bennett, S. E.; Mosbaugh, D. W. *Nucleic Acids Res.* **1990**, *18*, 7317-7322.

22) Bost, D. A.; Stoffel, S.; Landre, P.; Lawyer, F. C.; Akers, J.; Abramson, R. D.; Gefland, D. H. *FASEB J.* **1994**, *8*, A1395.

23) Lundberg, K. S.; Shoemaker, D. D.; Adams, M. W.; Short, J. M.; Sorge, J. A.; Mathur, E. J. *Gene* **1991**, *108*, 1-6.

24) Mathur, E. J.; Adams, M. W.; Callen, W. N.; Cline, J. M. *Nucleic Acids Res.* **1991**, *19*, 6952.

25) Uemori, T.; Ishino, Y.; Toh, H.; Asada, K.; kato, L. *Nucleic Acids Res.* **1993**, *21*.

26) Ruttimann, C.; Cotoras, M.; Zalvidar, J.; Vicuna, R. *Eur. J. Biochem.* **1985**, *149*, 41-45.

27) Carballeira, N.; Nazabal, M.; Brito, J.; Garcia, O. *BioTechniques* **1990**, *9*, 276-281.

28) Perler, F. B. *unpublished data*, 1994.

29) Perler, F. B.; Comb, D. G.; Jack, W. E.; Moran, L. S.; Qiang, B.; Kucera, R. B.; Benner, J.; Slatko, B. E.; Nwankwo, D. O.; Hempstead, S. K.; Carlow, C. K. S.; Jannaach, H. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *6+*, 5577-5581.

30) Komberg, A.; Baker, T. A. *DNA Replication*; Freeman: New York, 1992.

31) Hill, F.; Loakes, D.; Brown, D. M. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4258-4263.

Chapter 2
Synthesis of Pyrazole containing Nucleobase Analogs and Incorporation into DNA oligonucleotides

2.1 Synthetic Design of Pyrazole Nucleosides

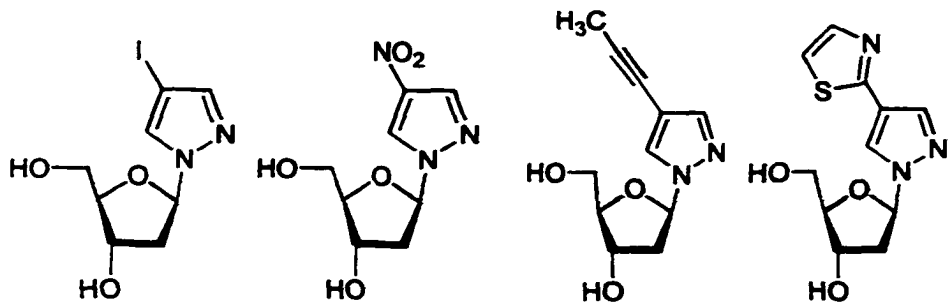
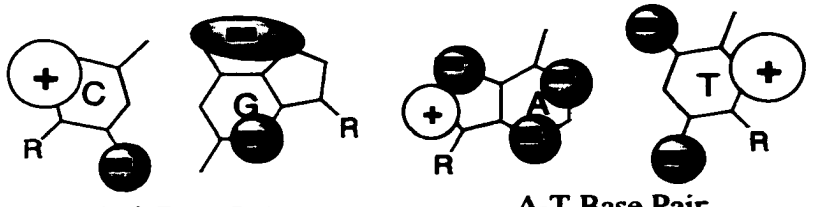


Figure 2.1 Target pyrazole nucleosides

For N-glycosides, our synthetic approach, similar to Bergstrom,[1] uses a chloro group as the leaving group.[1] Nucleophilic attack on the C1 position by N1 of the pyrazole displaces the chloro and results in the pyrazole nucleoside. Iodo and nitro substituents on the 4 position of pyrazole are available commercially. Palladium mediated chemistry is used to convert iodo into propyne and thiazole substituents. The propyne is directly converted using Pd catalyzed chemistry introduced by Robins.[2] The thiazole group is attached via a tin intermediate following Stille type coupling methods used by Froehler.[3] An initial approach to **1e** involved stannlyation of 2-bromothiazole using n-butyl lithium and tributyltinchloride, but attempts to isolate 2-tributyltin-thiazole away from the tin byproducts proved difficult by distillation.[4] Therefore the tin group was introduced

to the iodonucleoside and coupling of the thiazole to the pyrazole-nucleoside proved more effective.

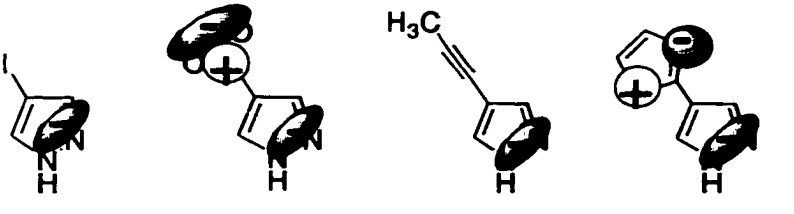
The natural nucleosides offer a charge distribution about the heterocyclic systems (Figure 2.2). The charge distribution not only offers electrostatic attractions across the base pairs but also between base pairs along the axis of the helix. The natural bases also offer large aromatic systems, which are capable of π-stacking, meaning favorable interactions between the large π–clouds along the center of the helix.



C-G Base Pair A-T Base Pair

Figure 2.2 Charge Distribution in Natural Base Pairs

The pyrazole analogs were designed to offer both charge distribution and the ability to stack. By varying the type of substituents, the importance of each contribution might be better understood. The thiazole ring offers a π-system for stacking as well as positive and negative centers for charge distribution. The propynyl unit offers π-stacking and the nitro unit offers a large negative center while the iodo group offers similar size but not the same charge or π-system (Figure 2.3).



PzI **PzN** **PzP** **PzT**

Figure 2.3 Charge Distribution of Pyrazole Analogs

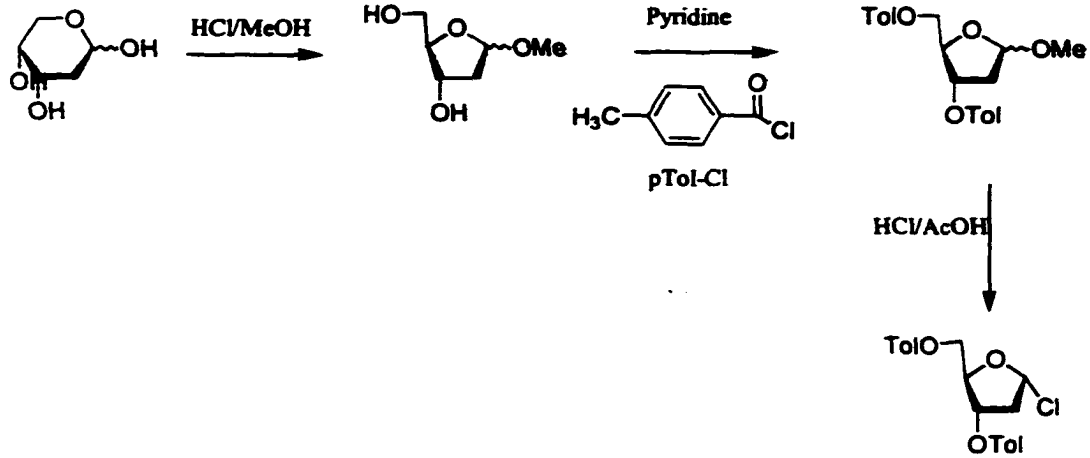
2.2 Synthesis of Nucleosides

4-Iodopyrazole and 4-nitropyrazole toluoyl protected nucleosides (**1a** and **1b**) were synthesized by treating the appropriate pyrazole (4-iodopyrazole or 4-nitropyrazole) with NaH followed by addition to 3,5-bis-*O*-*p*-toluoyl-β-D-2-deoxyribose chloride (Scheme2.1)[5] in 72% to 86% yield (Scheme 2.2). 4-Propynyl-pyrazole nucleoside was prepared from **1a** by treatment with propyne gas, $Pd(Cl_2)(PPh_3)_2$, triethylamine and copper iodide in 60% yield.[2] When water is not rigorously excluded, then a reduction of 4-iodopyrazole **1a** to an unsubstituted pyrazole nucleoside is observed (see experimental section for details). 4-(2-Thiazolyl) pyrazole nucleoside (**1e**) was also made from **1a** by conversion to 4-(trimethyltin) pyrazole tolouyl protected nucleoside (**1d**) (86%) followed by coupling with 2-bromothiazole using $Pd(PPh_3)_4$ (60%).[3]

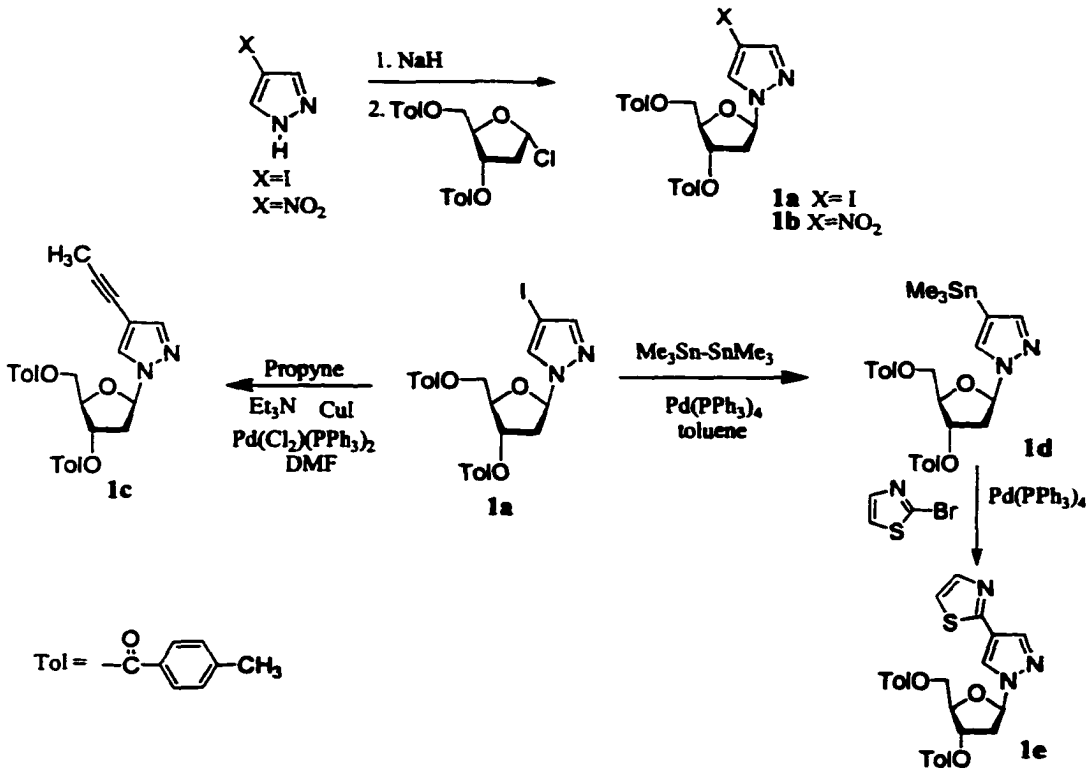
This final synthetic approach for **1e** deserves some further comment. As mention in Section 2.1, introduction of the tin group directly to the thiazole did not prove to be effective. Inverting the approach by introduction of the tin group to the nucleoside allowed purification of the tin intermediate **1d** by flash chromatography in excellent yield (86%). However, care must be taken to avoid protodestannylation of the nucleoside on the silica column, which again results in an unsubstituted pyrazole nucleoside, by the use of triethylamine (0.5%) in the eluant. Addition of the tin moiety to the nucleoside in our hands worked best in a non-polar solvent like toluene and the thiazole coupling required a polar solvent to provide any thiazolyl-pyrazole product (DMF yeilds 50-60%, dioxane yeilds 10-20%).

The free nucleosides, **2a, 2b, 2c, and 2e,** were synthesized by treatment of the toluoyl protected nucleosides with methanol saturated with ammonia. The free nucleosides were then prepared for incorporation into DNA oligomers by utilizing standard chemistry to achieve the dimethoxytrityl-protected phosphoramidite (**4a, 4b, 4c, 4e**) (Scheme 2.3).[6]

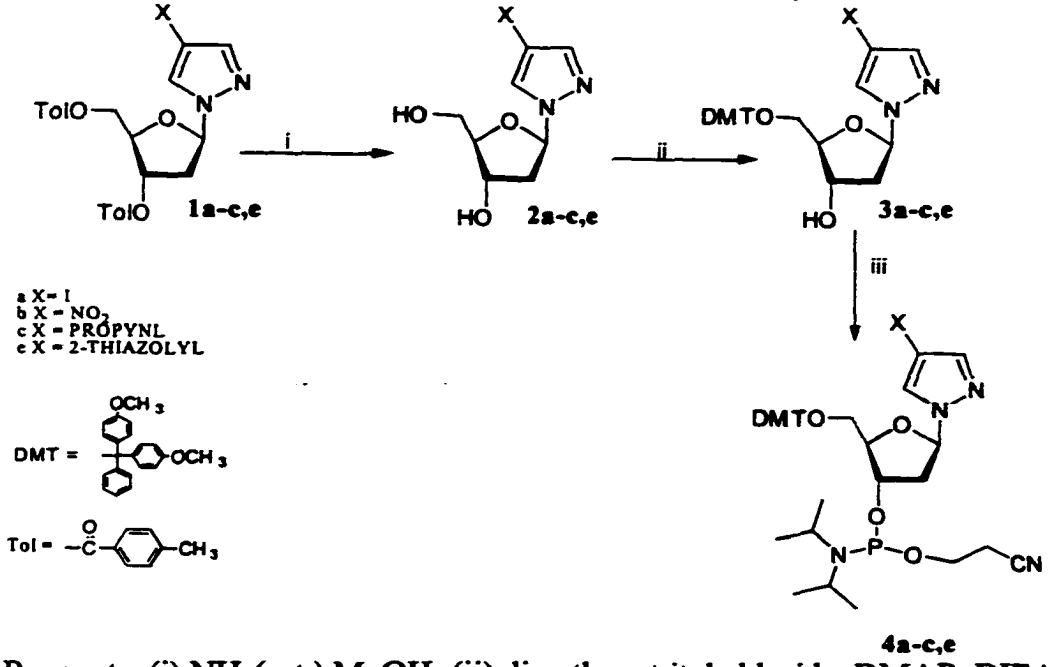
Scheme 2.1 Synthesis of 3,5-bis-*O*-*p*-toluoyl-β-D-2-deoxyribosyl chloride



Scheme 2.2 Synthesis of 4-substitued pyrazole protected nucleosides



Scheme 2.3 Synthesis of 4-substituted pyrazole phosphoramidite for incorporation into oligomers using automated DNA synthesis



Reagents: (i) NH_3(sat.) MeOH; (ii) dimethoxytrityl chloride, DMAP, DIEA, DCM; (iii) 2-cyanoethyl N,N-diisopropylchlorophosphoramidite, pyridine

2.3 Incorporation of Nucleosides Analogs into DNA Oligomers

2.3.1 Solid Phase DNA Synthesis

Solid phase DNA synthesis (Scheme 2.4) is the most practical and efficient method for incorporation of the modified bases. This method is rapid and simple. Utilizing the phosphoramidite method for synthesizing oligomers[7](Scheme 2.4), the reactive phosphorous (III)[8] on the 3' end of one base couples with the free 5' hydroxyl of a tethered nucleoside. Control Pore Glass (CPG)[9] beads are used as the solid support with the first base attached through a succinate ester linkage which is base labile. The natural bases are protected using base labile protecting groups such as benzoyl. (Figure 2.4) These protecting groups are cleaved once the synthesis is complete at 55 °C for 8 hours in ammonium hydroxide. The 5' hydroxyl of the incoming base is protected with dimethoxytrityl, an acid labile group.[6] The acid

labile protecting group is removed using mild acid. The next step requires the coupling of the activated 3'-P(III) center of the incoming base to the free 5'OH of the tethered using tetrazole as the activating species.[10] Once the coupling of the new base with the tethered base is complete, a capping step is performed using acetic anhydride to block any unreacted hydroxyl groups from reacting in later steps, thus avoiding the creation of deletion sequences. Once the capping step is complete. The P(III) center is oxidized to P(V) using aqueous iodine. This cycle is repeated n times until the desired sequence is synthesized. Cleavage of the oligomer off the support requires the used of ammonium hydroxide for periods of 15 minutes. Once removed from the support the base protecting groups are removed at 55 °C for 8 h.

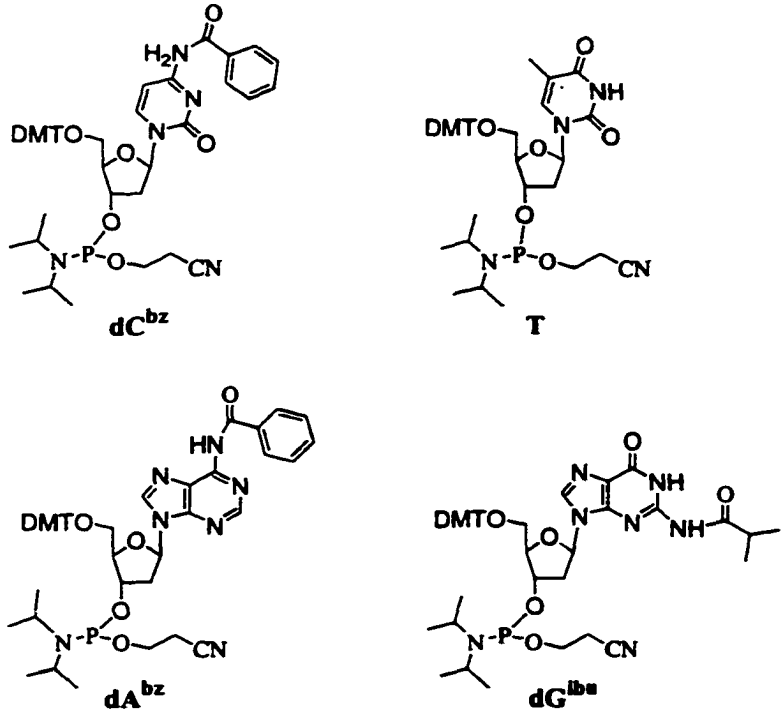
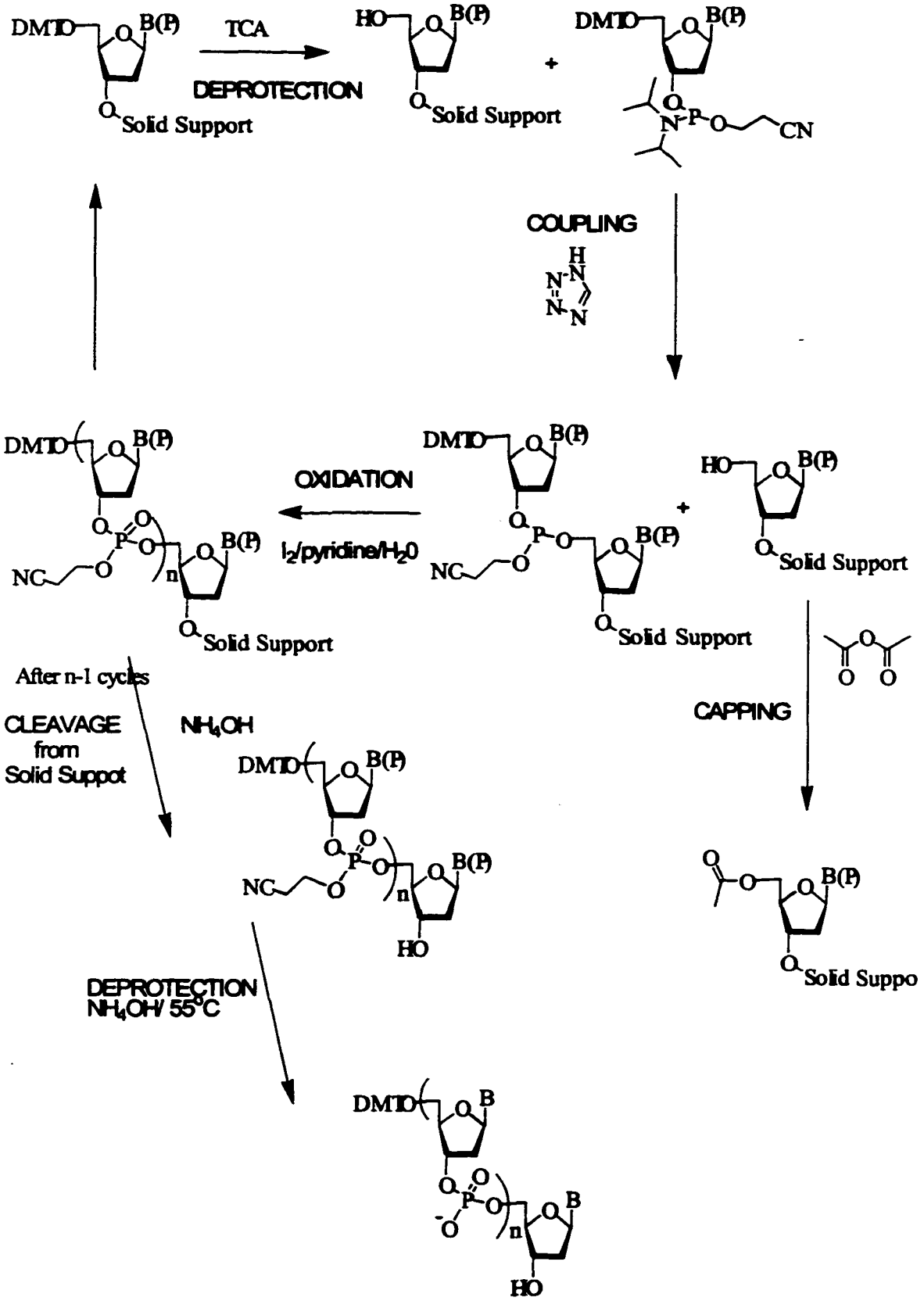


Figure 2.4 Phosphoramidites used in solid phase synthesis

Scheme 2.4 Solid phase DNA synthesis, phosphoramidite method



2.3.2 Preparation of Oligonucleotides with Analogs

Incorporation of the pyrazole analogs into DNA strands was done using standard automated DNA chemistry. (See Sections 3.3 and 4.7.1.3 for sequences synthesized.) Oligodeoxyribonucleotides were prepared from commercially available dA, dC, dG, dT (ABD-PE) and 5-nitroindole phosphoramidite (Glen Research) on an ABI 394 synthesizer using standard solid phase phosphoramidite chemistry (1 μmol scale). An alternative oxidation method (Chapter 6) was used with long sequences containing 5-nitroindole. Stepwise coupling yields ranged from 80 to 100%. The 5'-O-DMT protected oligonucleotides were synthesized and cleaved off the Control Pore Glass (CPG) column using ammonium hydroxide. The protecting groups for the natural bases were removed by ammonium cleavage at 55 °C for 8 hours. The crude reaction mixture was then purified using "trityl on" purification techniques with Oligopurification Cartridges (OPC) from ABI-Perkin-Elmer according to the manufacture's instructions. The eluant was checked for purity by reverse phase or anion exchange high performance liquid chromatography (HPLC). In some instances the oligomer was purified using preparative anion exchange HPLC. Anion Exchange HPLC was performed using a Hydrocell NS1000 polymeric column 4.5 x 150 mm on a Perkin Elmer Series 200 HPLC with 785A UV/VIS detector at 260 nm. Characterization was also by MALDI mass spectrometry (PerSeptive Biosystems Inc, Voyager Linear MALDI-TOF with N_2 laser and ammonium citrate matrix).

2.3.3 Calculation of Extinction Coefficient

The extinction coefficient for each of the unprotected nucleosides (nitro, iodo, propynyl and thiazolyl) was determined using a Beer's Law plot of the absorbance (260 nm) at each of 5 known concentrations. The extinction coefficient of 5-nitroindole is 15,296 M^{-1} cm^{-1}.[11]

Table 2.1 Extinction coefficient for each unprotected pyrazole nucleoside

Compound	Extinction coefficient M^{-1} cm^{-1}
Ind	15,296
(2a) PzI	285
(2b) PzN	5810
(2c) PzP	954
(2e) PzT	6927

2.4 Experimental

Compounds **2a, 3a, 4a, 1b, 2b, 3b, 4b, 2c, 3c,** and **4c** were synthesized by either Melissa Cameron or Tod Miller for incorporation into oligomers

2.4.1 (2'-Deoxy-3', 5'-bis-*O*-*p*-toluoyl-β-D-ribofuranosyl)-4-iodopyrazole (1a)

Procedure A: To an oven dried two neck flask containing 4-iodopyrazole (6.48 g, 33.4 mmol), acetonitrile (30 mL) was added under Ar by cannulation, cooled to 0 °C, followed by addition NaH (1.01 g, 42.1 mmol). Once the solution ceased bubbling, 3, 5-bis-*O*-*p*-toluoyl-β-D-2-deoxy-ribofuranosyl chloride[5] was added (13.0 g, 33.7 mmol) and the solution was stirred under Ar and allowed to warm to room temperature overnight. The solution was concentrated under reduced pressure

and purified by silica column (2:3 hexanes-CH_2Cl_2) to yield 13.3 g (72%) of a white solid. m.p. 110 °C; TLC: R_f= 0.66 (2:1 hexanes-EtOAc); FABMS (Glycerol) Theoretical 547.3 $(M+H)^+$, Found 547.3 $(M+H)^+$; ^1H (250 MHz, $CDCl_3$), 7.94 (d, J_{Ar}=4.1, 2H, Ar), 7.91 (d, J_{Ar}=4.2, 2H, Ar), 7.64 (s, 1H, H5), 7.54 (s, 1H, H3), 7.27 (d, J_{Ar}=2.9, 2H, Ar), 7.22 (d, J_{Ar}=7.8, 2H, Ar), 6.20 (t, J_{H2}=6.2, 1H, H1'), 5.75 (dd, J=2.8, J=5.9, 1H, H3'), 4.63-4.49 (m, 3H, H4 and H5), 3.20-3.09 (m, 1H, H2'), 2.66 (ddd, J=3.7, 6.2, 9.8, 1H, H2'), 2.42 (s, 3H, CH_3), 2.41 (s, 3H, CH_3); ^{13}C (75 MHz, $CDCl_3$) δ, 166.17 (C(O)), 165.82 (C(O)), 145.23 (C5), 144.26 (C(O)-Ar), 143.82 (C(O)-Ar), 133.19 (C3), 129.66 (Ar), 129.11 (Ar), 126.83 (Ar-CH_3), 126.55 (Ar-CH_3), 89.84 (C1'), 82.84 (C4'), 75.06 (C3'), 64.03 (C5'), 37.41 (C2'), 21.66 (CH_3); Anal. Calc'd for $C_{24}H_{23}IN_2O_5$: C 52.76, H 4.24, N 5.13,. Found: C 52.57, H 4.50, N 5.11.

2.4.2 (1-(2'-Deoxy-β-D-ribofuranosyl)-4-iodopyrazole (2a)

Procedure B: Deprotection of 1-(2'-deoxy-3',5'-bis-*O*-*p*-toluoyl-β-D-ribofuranosyl)-4-iodopyrazole (**1a**) (2.08 g, 3.82 mmol) using ammonia saturated methanol (100 mL) overnight. The solvent was removed under reduced pressure followed by flash chromatography (4:5 CH_2Cl_2-EtOAc) to yield a yellowish oil (1.01 g, 86%). TLC: R_f=0.39 (9:1 $CHCl_3$-CH_3OH); UV-Vis: ε_{260} =285 $M^{-1}cm^{-1}$; FABMS (Glycerol) Theoretical 311.1$(M+H)^+$, Found 311.3 $(M+H)^+$; ^1H (250 MHz, $CDCl_3$) δ 7.60 (s, 1H, H5), 7.59 (s, 1H, H3), 6.11 (t, J_{H2}=6.4, 1H, H1'), 4.74 (dd, J=2.7, 5.6, 1H, H3'), 4.13 (app. d, J=2.4, 1H, H4'), 3.88 (dd, J_{H4}=2.0, J_{H5}=12.6, 1H, H5'), 3.72 (dd, J_{H4}=2.0, J_{H5}=12.6, 1H, H5'), 2.89-2.78 (m, 1H, H2'), 2.41 (ddd, J=3.1, J=6.8, J=13.8, 1H, H2'). ^{13}C (62.5 MHz, $CDCl_3$) δ 145.78 (C5), 134.11 (C3), 89.92 (C1'),

89.04 (C4'), 72.72 (C3'), 63.36 (C5'), 41.64 (C2'); Anal. calc'd for $C_8H_{11}N_2O_3$: C 30.99, H 3.58, N 9.03, Found: C 30.75, H 3.80, N 9.05

2.4.3 1-(2'-Deoxy-5'-dimethoxytrityl-β- D-ribofuranosyl)-4-iodopyrazole (3a)

Procedure C: 1-(2'-Deoxy-β-D-ribofuranosyl)-4-iodopyrazole (**2a**) (0.999 g, 3.23 mmol) was coevaporated with 2 portions of pyridine then dissolved in pyridine (10 mL). Next addition of DMAP (0.424 g, 3.47 mmol) was followed by DIEA (8 mL) and DMT-Cl (1.05 g, 3.09 mmol). The reaction was left stirring overnight under inert atmosphere. Purification of the crude product by flash chromatography yielded 1.45 g (73%) of oil. m.p. = 35 °C; TLC: R_f=0.34 (4:3:1 hexanes-EtOAc-Et_3N); FABMS (NBA) 612.6 $(M)^+$; ^1H (200 MHz, $CDCl_3$) δ 7.64 (s, 1H, H5), 7.46 (s, 1H, H3), 7.40 (app. d, J=8.0, 2H, Ar), 7.24-7.16 (m, 7H, Ar), 6.80 (app. d, J=8.2, 4H, Ar), 6.09 (app. t, J_{H2}=5.1, 1H, H1'), 4.62-4.57 (m, 1H, H3'), 4.19 (app. dd, J=4.2, J=8.8, 1H, H4'), 3.78 (s, 6H, OCH_3), 3.34-3.17 (m, 2H, H5' & H5"), 2.79-2.67 (m, 1H, H2'), 2.57-2.438 (m, 1H, H2"); Anal. calc'd for $C_{29}H_{29}IN_2O_5$: C 56.87, H 4.77, N 4.57, Found: C 56.71, H 4.89, N 4.61

2.4.4 1-(2'-Deoxy-5'-dimethoxytrityl-3'-*O*-2-cyanoethyl-N,N-diisopropylphosphoramidityl-β-D-ribofuranosyl)-4-iodopyrazole (4a)

Procedure D: 1-(2'-Deoxy-5'-dimethoxytrityl-β-D-ribofuranosyl)-4-iodopyrazole (**3a**) (0.39 g, 0.638 mmol) was twice co-evaporated with pyridine then dissolved in CH_2Cl_2 (5 mL) and triethylamine (0.4 mL) under Ar. Next, diisopropylchlorocyanoethylphosphoramidite (150 μL, 0.67 mmol) was added slowly and the solution was stirred for 1 h. The crude reaction was quenched with methanol and then washed with brine, dried with Na_2SO_4 and concentrated under reduced pressure. Purification by flash chromatography (2:1:0.05 hexanes-EtOAc-

Et_3N) provided a white foam (0.366 g, 71%). TLC: R_f=0.35 (4:3:1 hexanes-EtOAc-Et_3N); HR FAB (MNBA) Theoretical: 813.2278 $(M+H)^+$, Found: 813.2262 $(M+H)^+$; ^{31}P (101.2 MHz, $CDCl_3$) mixture of diastereomers δ 147.62, 147.35. ^1H mixture of diastereomers; ^1H (250 MHz, $CDCl_3$) δ 8.23 (s, 1H, H5), 7.68 (s, 1H, H3), 7.42-7.39 (m, 2H, Ar), 7.31-7.19 (m, 7H, Ar), 6.82-6.78 (m, 4H, Ar), 6.05 (app. t, $J_{H2'}$=5.9, 1H, H1'), 4.69-4.63 (m, 1H, H4'), 4.20-4.16 (m, 1H, H3'), 3.78 (s, 6H, OMe), 3.66-3.48 (m, 4H, CH_3-**CH** and O-**CH_2**), 3.24-3.21 (m, 2H, H5'), 2.78-2.40 (m, 4H, C2 and **CH_2**-CN), 1.99 (s, 3H, CH_3),1.32-1.12 (m, 12H, CH_3-CH).

2.4.5 1-(2'-Deoxy-3',5'-bis-*O*-*p*-toluoyl-β-D-ribofuranosyl)-4-nitropyrazole (1b)

Prepared using procedure A with 4-nitropyrazole (2.77 g, 24.3 mmol), acetonitrile (20 mL), NaH (0.62 g, 25.8 mmol) in dry acetonitrile (10 mL), 3, 5-bis-*O*-*p*-toluoyl-β-D-2-deoxyribofuranosyl chloride (11.94 g, 30.7 mmol), purified by silica column ($CHCl_3$) to yield 12.24 g (86%) of a white solid. m.p. 157-158 °C; TLC: R_f=0.50 (2:1 hexanes-EtOAc); HR FAB (MNBA) Theoretical: 466.1614 $(M+H)^+$, Found: 466.1606 $(M+H)^+$; ^1H (200 MHz, $CDCl_3$) δ 8.36 (s, 1H, H-5), 8.06 (s,1H, H-3), 7.94 (d, J_{Ar}=8.2, 2H, Ar), 7.88 (d, J_{Ar}=8.3, 2H, Ar), 7.27 (d, J_{Ar}=8.1, 2H, Ar), 7.22 (d, J_{Ar}=8.3, 2H, Ar), 6.20 (app.t, $J_{H2'}$=6.0, 1H, H1'), 5.77 (app. dt, $J_{H2'}$=3.2, $J_{H4'}$=6.3, 1H, H3'), 4.70-4.63 (m, 2H, H4' & H5'), 4.54 (dd, $J_{H4'}$=6.2, $J_{H5'}$=13.5, 1H, H5"), 3.22 (app. dt, J=6.2, $J_{H2''}$=14.2, 1H, H2'), 2.85 (ddd, $J_{H3'}$=4.0, $J_{H1'}$=6.4, $J_{H2''}$=14.4, 1H, H2"), 2.44, 2.41 (s, 3H, **CH_3**-Ar); ^{13}C (50 MHz, $CDCl_3$) 166.10, 165.80 (C(O)-Ar), 144.48, 144.11 (C(O)-Ar), 136.22 (C5), 129.65, 129.56, 129.16, 129.06, (Ar), 127.86, (C3), 126.58, 126.33 (Ar-CH_3), 105.00 (C4), 90.66 (C1'), 83.60 (C4'), 74.56 (C3'), 63.61 (C5'), 37.80 (C2'), 21.60, 21.55 (Ar-**CH_3**)

2.4.6 1-(2'-Deoxy-β-D-ribofuranosyl)-4-nitropyrazole (2b)

Prepared using procedure B with 1-(2'-deoxy-3',5'-bis-*O*-*p*-toluoyl-β-D-ribofuranosyl)-4-nitropyrazole (**1b**) (1.65 g, 3.54 mmol), ammonia saturated methanol (100 mL), purified by column chromatography with silica (4:5 CH_2Cl_2-EtOAc) to give a yellowish oil (0.717 g, 88%). TLC: R_f=0.17 (1:2 CH_2Cl_2-EtOAc); UV-Vis: ε_{260} =5818 $M^{-1}cm^{-1}$; HR FAB (MNBA) Theoretical: 230.0776 $(M+H)^+$, Found: 230.0756 $(M+H)^+$; ^1H NMR (200 MHz, CD_3OD) δ 8.76 (s, 1H, H5), 8.12 (s, 1H, H1), 6.14 (app. t, $J_{H2'}$=6.4, 1H, H1'), 4.50 (app. dd, $J_{H2'}$=5.5, $J_{H4'}$=10.5, 1H, H3'), 4.02 (app. dd, $J_{H5''}$=4.3, $J_{H3'}$=8.7, 1H, H4'), 3.75 (dd, $J_{H4'}$=3.7, $J_{H5''}$=11.9, 1H, H5''), 3.64 (dd, $J_{H4'}$=5.17, $J_{H5'}$=12.0, 1H, H5'), 2.84-2.71 (m, 1H, H2'), 2.52 (ddd, $J_{H3'}$=1.7, $J_{H1'}$=7.1, $J_{H2''}$=13.9, 1H, H2'); ^{13}C (250 NMR, CD_3OD) δ 137.01 (C5), 130.01 (C3), 92.15 (C1'), 89.70 (C4'), 72.17 (C3'), 63.35 (C5'), 41.44 (C2').

2.4.7 1-(2'-Deoxy-5'-dimethoxytrityl-β-D-ribofuranosyl)-4-nitropyrazole (3b)

Prepare using procedure C with 1-(2'-deoxy-β-D-ribofuranosyl)-4-nitropyrazole (**2b**) (0.717 g, 3.13 mmol) dry pyridine (15 mL), DIEA (10 mL), DMT-Cl (1.08 g, 3.18 mmol) and DMAP (0.38 g, 3.11 mmol), purified by column chromatography with silica (20:1 EtOAc-CH_3OH) to give 1.47 g (88%) of oil. TLC: R_f = 0.18 (4:3:1 hexanes-EtOAc-Et_3N); FAB MS (NBA) Theoretical 531.5 $(M)^+$, Found 531.4 $(M)^+$; ^1H (250 MHz, CD_3OD) δ 8.60 (s, 1H, H5), 8.03 (s, 1H, H3), 7.38-7.35 (dd, J=7, 2H, Ar), 7.27-7.11 (m, 7H, Ar), 6.77-6.73 (d, J=8.5, 4H, Ar), 6.12-6.08 (dd, J=4, J=2.3 1H, H1'), 6.12-6.08 (dd, J=5.01H, H3'), 4.58-4.56 (dd, 1H, H4'), 3.69 (s, 6H, OMe), 4.13-4.03 (m, 2H, H5' & H5''), 2.82-2.77 (m, 1H, H2'), 2.46-2.36 (m, 1H, H2''); ^{13}C (300 MHz, CD_3OD) δ 160.12 (Ar-OMe), 146.38 (C5), 137.25, 137.17 (Ar-C),

136.96 (C3), 131.35, 129.99, 129.38, 128.83, 127.87 (Ar), 114.15 (C4), 91.98 (C1'), 88.35 (C4'), 87.62 (C-Ar), 72.43 (C3'), 65.21 (C5'), 55.79 (OMe), 40.91 (C2').

2.4.8 1-(2'-Deoxy-5'-dimethoxytrityl-3'-*O*-2-cyanoethyl-N,N-diisopropylphosphoramidityl-β-D-ribofuranosyl)-4-nitropyrazole (4b)

Prepared using procedure D with 1-(2'-deoxy-5'-dimethoxytrityl-β-D-ribofuranosyl)-4-nitropyrazole (**3b**) (0.52 g, 0.9787 mmol), CH_2Cl_2 (5 mL), triethylamine (0.75 mL), diisopropylchlorocyanoethylphosphoramidite (0.3 mL, 1.34 mmol) and purified by column chromatography with silica (2:1:0.05 hexanes-EtOAc-Et_3N) to give a white foam (0.622 g, 87%). TLC: $R_f = 0.70$ (4:3:1 hexanes-EtOAc-Et_3N); HR FAB (MNBA) Theoretical: 732.3161 $(M+H)^+$, Found: 732.3153 $(M+H)^+$; ^{31}P mixture of diastereomers (101.2 MHz, $CDCl_3$) δ 147.81, 147.63. ^1H mixture of diastereomers (250 MHz, $CDCl_3$) δ 8.37, 8.35 (s, 1H, H5), 8.04 (s, 1H, H3), 7.39-7.36 (m, 2H, Ar), 7.29-7.20 (m, 7H, Ar), 6.82-6.78 (m, 4H, Ar), 6.08 (dd, $J_{H2'}$=5.8, $J_{H2''}$=8.7, 1H, H1'), 4.68 (m, 1H, H4'), 4.28 (app. t, J=4.2, 1H, H3'), 3.78 (s, 6H, OMe), 3.70-3.52 (m, 4H, CH_3-**CH** and O-**CH_2**), 3.26 (dd, J=7.4, J=14.3, 2H, H5'and H5''), 2.84-2.56 (m, 4H, C2 and **CH_2**-CN), 1.26-1.08 (m, 12H, CH_3-CH).

2.4.9 1-(2'-Deoxy-3',5'-bis-*O*-*p*-toluoyl-β-D-ribofuranosyl)-4-propynylpyrazole (1c)

1-(2'-Deoxy-3',5'-bis-*O*-*p*-toluoyl-β-D-ribofuranosyl)-4-iodopyrazole (**1b**) (0.545 g, 0.998 mmol) was dissolved in DMF (5 mL) and triethylamine (1 mL, 7.5 mmol) into a 90 mL pressure vessel. The solution was purged with Ar and cooled to –78 °C. 2 mL of propyne (33 mmol) was condensed into the vessel followed by addition of CuI (0.053 g, 2.63 mmol) and $Pd(PPh_3)_2Cl_2$ (0.072 g, 0.103 mmol). The flask was sealed and warmed to room temperature with stirring for 24 hours. The

solvents were removed under reduced pressure and the crude product was purified using flash chromatography with silica ($CHCl_3$) to give 0.400 g (0.873mmol) of a white solid (87%). m.p. 110-112 °C; TLC: R_f=0.22 (3:1 ether-hexanes); HR FAB (MNBA) Theoretical 459.1920 $(M+H)^+$, Found 459.1920 $(M+H)^+$; ^1H (250 MHz, $CDCl_3$) δ 7.93 (d, J_{Ar}=7.9, 4H, Ar), 7.65 (s, 1H, H5), 7.57 (s, 1H, H3), 7.27 (d, J_{Ar}=2.4, 2H, Ar), 7.21 (d, J_{Ar}=7.7, 2H, Ar), 6.16 (t, $J_{H2'}$=6.1, 1H, H1'), 5.76 (app. dd, $J_{H2'}$=3.2, $J_{H4'}$=6.4, 1H, H3'), 4.60-4.48 (m, 3H, H4' and H5'), 3.21-3.11 (m, 1H, H2'), 2.66 (ddd, $J_{H1'}$=6.6, $J_{H3'}$=3.7, $J_{H2'}$=14.4, 1H, H2'), 2.42 (s, 3H, Ar-**CH_3**), 2.40 (s, 3H, Ar-**CH_3**), 1.99 (s, 3H, C-**CH_3**); ^{13}C (50 MHz, $CDCl_3$) δ 166.25 (C(O)), 165.87 (C(O)), 144.26 (C(O)-**Ar**), 143.78 (C(O)-**Ar**), 142.75 (C5), 131.10 (C3), 129.78 (Ar), 129.22 (Ar), 129.12 (Ar), 129.02 (Ar), 126.98 (**Ar**-CH_3), 126.67 (**Ar**-CH_3), 104.83 (C4), 89.78 (C1'), 86.52 (Pyrazole-**C**), 82.74 (C4'), 75.21 (C3'), 70.24 (**C**-CH_3), 64.21 (C5'), 37.35 (C2'), 21.69 (Ar-**CH_3**), 4.27 (C-**CH_3**); Anal. calc'd for $C_{27}H_{26}N_2O_5$: C 70.73, H 5.72, N 6.11, Found: C 70.56, H 5.71, N 5.92

2.4.10 1-(2'-Deoxy-β-D-ribofuranosyl)-4-propynylpyrazole (2c)

Prepared using procedure B with 1-(2'-deoxy-3',5'-bis-*O*-*p*-toluoyl-β-D-ribofuranosyl)-4-propynylpyrazole (**1c**) (1.0 g, 2.18 mmol), ammonia saturated methanol (100 mL) and purified using flash chromatography (1:2 CH_2Cl_2-EtOAc) to give a yellowish oil.(0.41 g, 85%). TLC: R_f=0.24 (9:1 $CHCl_3$-CH_3OH); UV-Vis: ε_{260} =127 $M^{-1}cm^{-1}$; HR FAB (MNBA) Theoretical 223.1083 $(M+H)^+$, Found 223.1082 $(M+H)^+$; ^1H (250 MHz, $CDCl_3$) δ 7.60 (s, 1H, H5), 7.56 (s, 1H, H3), 6.00 (t, $J_{H2'}$=6.2, 1H, H1'), 5.09 (app. d, J=5.4, 1H, H3'), 4.04 (app. d, J=2.3, 1H, H4'), 3.81-3.77 (m, 1H, H5'), 3.69-3.61 (m, 1H, H5'), 2.74-2.63 (m, 1H, H2'), 2.32 (ddd,

J=1.5, J=5.8, J=13.0, 1H, H2'), 1.96 (s, 3H, CH_3); ^{13}C (62.5 MHz, $CDCl_3$) δ 143.00 (C5), 131.68 (C3), 104.04 (C4), 89.79 (C1'), 88.87 (C4'), 86.69 (pyrazole-C), 72.44 (C3'), 69.76 (C-CH_3), 63.25 (C5'), 41.50 (C2'), 4.15 (C-CH_3); Anal. Calc'd for $C_{11}H_{14}N_2O_3$: C 59.45, H 6.35, N 12.61, Found: C 59.27, H 6.34, N 12.36.

2.4.11 1-(2'-Deoxy-5'-dimethoxytrityl-β-D-ribofuranosyl)-4-propynylpyrazole (3c)

Prepared using procedure C with 1-(2'-deoxy-β-D-ribofuranosyl)-4-propynylpyrazole (2c)(0.41 g, 1.845 mmol), DMT-Cl (0.938 g, 2.77 mmol) and DMAP (0.338 g, 2.76 mmol) and purified using flash chromatography (4:3:1 hexanes-EtOAc-Et_3N) to give 0.757 g (83%) of oil. TLC: R_f=0.54 (4:3:1 hexanes-EtOAc-Et_3N); FABMS (glycerol) Theoretical 525.6 $(M+H)^+$, Found 525.5$(M+H)^+$; ^1H (200 MHz, $CDCl_3$) δ 7.54 (s, 1H, H5), 7.41 (s, 1H, H3), 7.34-7.31 (m, 2H, DMT), 7.23-7.07 (m, 7H, DMT), 6.75-6.69 (m, 4H, DMT), 5.93 (dd, $J_{H2'}$=4.6, $J_{H2'}$=6.6, 1H, H1'), 4.50 (dd, J=5.8, J=10.9, 1H, H3'), 3.92 (dd, J=5.1, J=10.1, 1H, H4'), 3.71 (s, 6H, OCH_3), 3.27-3.09 (m, 2H, H5'), 2.73-2.61 (m, 2H, H2'), 1.91 (s, 3H, CH_3); ^{13}C (50 MHz, $CDCl_3$) δ 158.48 (Ar-OCH_3), 149.10, 144.67 (DMT), 142.24 (C5), 135.84 (C3), 130.73, 130.02, 128.13, 127.82, 126.76, 113.14 (DMT), 104.39 (C4), 89.24 (C1'), 86.40 (C-$(Ar)_3$), 85.63 (Pyrazole-C), 85.63 (C4'), 72.37 (C3'), 70.42 (C-CH_3), 64.06 (C5'), 55.15 (OCH_3), 39.73 (C2), 4.21 (C-CH_3); Anal. Calc'd for $C_{32}H_{32}N_2O_5 \bullet H_2O$: C 70.83, H 6.32, N 5.16, Found: C 69.95, H6.08, N5.12.

2.4.12 1-(2'-Deoxy-5'-dimethoxytrityl-3'-*O*-2-cyanoethyl-N,N-diisopropylphosphoramidityl-β-D-ribofuranosyl)-4-propynylpyrazole (4c)

Prepared using procedure D with 1-(2'-deoxy-5'-dimethoxytrityl-β-D-ribofuranosyl)-4-propynylpyrazole (**3c**) (1.093 g, 2.22 mmol), diisopropylchlorocyanoethylphosphoramidite (0.3 mL, 1.34 mmol), purified using column chromatography with silica (2:1:0.05 hexanes-EtOAc-Et_3N) to give a white foam (1.313 g, 89%). TLC: R_f= 0.56 (4:3:1 hexanes-EtOAc-Et_3N); HR FAB (MNBA) Theoretical: 725.3479 $(M+H)^+$, Found: 725.3427 $(M+H)^+$; ^{31}P (101.2 MHz, $CDCl_3$) mixture of diasteromers δ 155.28, 155.01; ^1H mixture of diastereomers (250 MHz, $CDCl_3$) δ 7.67 (s, 1H, H5), 7.50 (s, 1H, H3), 7.43-7.40 (m, 2H, Ar), 7.31-7.19 (m, 7H, Ar), 6.82-6.78 (m, 4H, Ar), 6.05 (app. t, $J_{H2'}$=5.9, 1H, H1'), 4.69-4.63 (m, 1H, H4'), 4.20-4.16 (m, 1H, H3'), 3.78 (s, 6H, OCH_3), 3.66-3.48 (m, 4H, CH_3-**CH** and O-**CH_2**), 3.24-3.21 (m, 2H, H5'), 2.78-2.40 (m, 4H, C2 and **CH_2**-CN), 1.99 (s, 3H, CH_3),1.32-1.12 (m, 12H, CH_3-CH).

2.4.13 1-(β-D-2'-Deoxyribofuranosyl)-3',5'-bis-*O*-(*p*-toluoyl)-4-trimethylstannyl-pyrazole (1d)

To a solution of 1-(β-D-2'-deoxyribofuranosyl)-4-iodo-pyrazole (**1a**) (3.21 g, 5.88 mmol) and toluene (70 mL) under N_2 was added hexamethyldistannane (2.7 mL, 13.0 mmol) followed by rapid addition of $Pd(PPh_3)_4$ (60 mg, 5.19 μmol). The reaction mixture was stirred and heated to 90 °C for 3 h and then concentrated under reduced pressure. The crude product was purified by flash chromatography (8:2:1 hexanes-EtOAc-Et_3N) providing a clear oil (2.90 g, 86%). TLC: R_f 0.42 (4:1, hexanes-EtOAc); HR FAB (MNBA) Theoretical 585.1411$(M+H)^+$, Found 585.1385

$(M+H)^+$; ^1H (200 MHz, $CDCl_3$) δ 8.06 (d, 4H, Ar,), 7.68 (s, 2H, pyrH), δ 7.29 (dd, *J*=8.0 Hz, 4H, Ar,), 6.4 (t, *J*=7.9 Hz, 1H, H1′), 5.90 (br s, 1H, H3′), 4.60-4.74 (m, 3H, H4′/H5′), 3.31 (dt, $J_{H2'}$=13.7 Hz, *J*= 6.7 Hz, 1H, H2′), 2.76 (ddd, $J_{H2'}$=13.7, *J* = 6.1, 3.1 Hz, 1H, H2′), 2.47 (s,3H, Ar-CH_3), 2.45 (s, 3H, Ar-CH_3), 0.32 (s, 9H $(CH_3)_3$-Sn; $^{117}Sn/^{119}Sn$ satellites, 2d, *J*= 56, 46 Hz); ^{13}C (50 MHz, $CDCl_3$) δ 165.97 (Ar-C(O)), 165.65 (Ar-C(O)), 145.91 (C5; $^{117}Sn/^{119}$ Sn satellite, 2d, *J*= approx. 443 Hz), 143.92 ((Ar-C(O)), 143.48 (Ar-C(O)), 133.15 (C3; $^{117}Sn/^{119}Sn$ satellite, 2d, *J*= approx. 66 Hz), 128.59 (Ar), 129.52 (Ar), 129.30 (Ar), 128.96 (Ar), 126.83 (Ar-CH_3), 126.57 (Ar-CH_3). 111.84 (C4; ^{117}Sn ^{119}Sn satellite, 2d, *J*= approx. 443 Hz), 89.27 (C1′), 82.33 (C3′), 75.27 (C4′), 64.21 (C5′), 37.37 (C2′), 21.43 (Ar-CH_3), -9.43 ($(CH_3)_3$-Sn); ($^{117}Sn/^{119}Sn$ satellites, 2d, *J*=352, 371 Hz); Anal. calc'd for $C_{27}H_{32}N_2O_5Sn$: C, 55.47.; H, 5.52; N, 4.79; Found C, 55.60: H, 5.33: N, 4.78.

2.4.14 1-(β-D -2′-Deoxyribofuranosyl)-3′,5′-bis-*O*-(*p*-toluoyl)-4-(2-thiazolyl)-pyrazole (1e)

To a solution of 1-(β-D-2′-deoxyribofuranosyl)-4-trimethylstannyl-pyrazole (**1d**) (2.90 g, 4.97 mmol) in DMF (62 mL) under N_2 was added 2-bromothiazole (2.77 mL, 30.4 mmol) followed by rapid addition of $Pd(PPh_3)_4$ (0.055 g, 47.5 μmol). The reaction mixture was stirred and heated to 90 °C for 20 h and then concentrated under reduced pressure. The crude product was purified by flash chromatography (2:1 hexanes-EtOAc) providing a white solid (1.49 g, 60%).TLC: R_f 0.38 (2:1 hexanes-EtOAc); m.p. 134-136 °C; HR FAB (MNBA) Theoretical 504.1595 $(M+H)^+$, Found 504.1624 $(M+H)^+$; ^1H (200 MHz $CDCl_3$) δ 8.10 (s, 1H, thiazole H4), 7.90-7.97 (m, 5H, Ar, thiazole H5), 7.75 (d, *J*=3.5 Hz, 1H, H5), δ 7.16-7.28

(m, 5H, Ar, H3), 6.25 ppm (t, J=6.2 Hz, 1H, H1′), 5.82 (dt, J=6.2, 3.1 Hz, 1H, H3′), 4.49-4.68 (m, 3H, H4′/H5′), 3.19 (dt, $J_{H2'}$=13.8, J = 6.7 Hz, 1H, H2′), 2.73 (ddd, J_{H2} = 13.8, J = 7.1, 3.5 Hz, 1H, H2′), 2.43 (s, 3H, Ar-**CH$_3$**), 2.38 (s, 3H, Ar-**CH$_3$**); ^{13}C (50MHz, $CDCl_3$) δ 165.62 (Ar-**C(O)**), 165.25, (Ar-**C(O)**), 159.52 (thiazole C2), 143.16 (pyrazole C5), 143.68 (**Ar**-C(O)), 142.39 (**Ar**-C(O)), 138.08 (pyrazole C3), 129.17 (Ar), 128.61 (Ar), 128.46 (Ar), 126.7 (thiazolyl C5), 126.33 (**Ar**-CH_3), 126.07 (**Ar**-CH_3), 117.84 (pyrazole C4), 116.5 (thiazolyl C4), 89.47 (C1′), 82.41 (C3′), 74.56 (C4′), 63.52 (C5′), 37.01 (C2′), 21.07 (Ar-**CH$_3$**), 21.00 (Ar-**CH$_3$**); Anal. calc'd for $C_{27}H_{25}N_3O_5S$: C, 64.40; H, 5.00; N, 8.31; found C, 64.27: H, 5.33: N, 8.31

2.4.15 1-(β-D -2′-Deoxyribofuranosyl)-4-(2-thiazolyl)-pyrazole (2e)

Prepare using procedure B with1-(β-D-2′-deoxyribofuranosyl)-3′,5′-bis-*O*-(*p*-toluoyl)-4-(2-thiazolyl)-pyrazole (**1e**) (1.26 g, 2.43), ammonia saturated methanol (120 mL) and purified using flash chromatography (30:1 $CHCl_3$-CH_3OH) yielded a white solid (0.50 g, 76%). TLC: R_f 0.16 ($CHCl_3$); UV-Vis: ε_{260}=6,927 $M^{-1}cm^{-1}$; m.p. 89-91 °C; HR FAB (glycerol) Theoretical 268.0746 $(M+H)^+$, Found 268.0759 $(M+H)^+$; ^{1}H (250MHz, $CDCl_3$) δ 8.10 (s, 1H, thiazolyl H4), 7.95 (s, 1H, thiazolyl H5), 7.72 (d, J=3.3 Hz, 1H, H5), 7.21 (d, J=2.85, 1H, H3), 6.17 (t, J=6.2 Hz, 1H, H1′), 4.91 (br s, 1H, OH3′), 4.74 (s, 1H, H3′), 4.26 (br s, 1H, OH5′), 4.16 (s, 1H, H5′), 3.80 (dd, J=12.2, 11.7, 2H, H4′) 2.83 (dt, J_{H2}=13.1, J = 6.4 Hz, 1H, H2′), 2.45 (ddd, J_{H2} = 13.6, J = 6.5, 3.4 Hz, 1H, H2′); ^{13}C (50MHz, $CDCl_3$) δ 159.52 (thiazole C2), 142.3 (pyrazole C5), 138.08 (pyrazole C3), 127.5 (thiazolyl C5), 117.0

(pyrazole C4), 116.9 (thiazolyl C4), 89.7 (C1′), 88.4 (C3′), 71.7 (C4′), 62.7 (C5′), 41.0 (C2′); Anal. calc'd for $C_{11}H_{14}N_3O_3S$: C, 49.43; H, 4.91; N, 15.73; Found C, 49.32: H, 5.11; N, 15.53.

2.4.16 (3e) 1-[5′-*O*-(Dimethoxytrityl -β-D -2′-deoxyribofuranosyl)]-4-(2-thiazolyl)-pyrazole

Prepared using procedure C with 1-(β-D -2′-deoxyribofuranosyl)-4-(2-thiazolyl)-pyrazole (**2e**) (0.409 g, 1.53 mmol) , pyridine (10 mL), DMAP (0.28 g 2.29 mmol), DIEA (3 mL) and DMT-Cl (0.789 g 2.32 mmol). Purification of the crude product by flash chromatography (200:2:1 $CHCl_3$-CH_3OH-Et_3N) yielded a yellowish solid (0.400 g, 46%). TLC: R_f 0.33 (9:1 $CHCl_3/CH_3OH$); m.p. 63-65 °C; HR FAB (MNBA) Theoretical 570.2062 $(M+H)^+$, Found 570.2064 $(M+H)^+$; ^1H (300MHz $CDCl_3$) δ 8.07 (s, 1H, thiazolyl H4), 7.91 (s, 1H, thiazolyl H5), 7.73 (s, 1H, H5), 7.41-7.17 (m. 9H, Ar-OCH_3, Ar-OCH_3, H3), 6.78 (d, 5H, Ar), 6.11 (dd, $J_{H2'}$=6.2, 4.6 Hz, 1H, H1′), 4.65 (app. dd, $J_{H4'}$=10.5 Hz, $J_{H2'}$=10.5 Hz, $J_{H2'}$=6.0 Hz,1H, H3′), 4.05 (app. dd, $J_{H3'}$=10.5 Hz, $J_{H5'}$=10.5 Hz, $J_{H5'}$=5.1 Hz,1H, H4′), 3.36 (app. dd, $J_{H4'}$=10.5 Hz, $J_{H5'}$=10.5 Hz, $J_{H5'}$=5.5 Hz,1H, H5′), 3.25 (app. dd, $J_{H5'}$=10.5 Hz, $J_{H4'}$=10.5 Hz, $J_{H2'}$=5.1 Hz,1H, H5″), 3.76(s, 6H, Ar-OCH_3, Ar-OCH_3) 2.85 (ddd $J_{H2'}$=12.3 Hz, $J_{H3'}$=10.5 Hz, $J_{H1'}$= 6.4 Hz, 1H, H2′); ^{13}C (75MHz, $CDCl_3$) δ 158.8 (pyrazole C5), 145.0 (thiazolyl C2), 138.6 (pyrazole C3), 127.5 (thiazolyl C4), 136.2 (pyrazole C4), 130.4, 128.5, 128.2, 127.5, 127.2 (DMT), 117.5 (thiazolyl C5), 113.5 (DMT), 89.95 (C1′), 86.8 DMT, 86.1 (C3′), 72.9 (C4′), 64.4 (C5′), 55.5(DMT), 40.3 (C2′); Anal. calc'd for $C_{32}H_{31}N_3O_5S$: C, 67.46; H, 5.49; N, 7.38; Found C,67.28; H,5.29 ; N, 7.35.

2.4.17 1-(2'-Deoxy-5'-dimethoxytrityl-3'-*O*-2-cyanoethyl-N,N-diisopropylphosphoramidityl-β-D-ribofuranosyl)-4-(2-thiazolyl)-pyrazole (4e)

Prepared using procedure D with 1-(2'-deoxy-5'-dimethoxytrityl-β- D - ribofuranosyl)-4-(2-thiazolyl)-pyrazole (**3e**) (150 mg, 195 μmol), CH_2Cl_2 (2 mL), triethylamine (1.5 mL), diisopropylchlorocyanoethylphosphoramidite (100 μL, 0.44 mmol). Purification by flash chromatography (2:1:1 hexanes/EtOAc/Et_3N) provided a yellowish oil (55 mg, 25%). TLC: R_f 0.25, 0.18 (2:1 hexanes-EtOAc); HR FAB (MNBA) Theoretical 770.3140 $(M+H)^+$, Found 770.3182 $(M+H)^+$; ^{31}P (101.2 MHz) δ 147.7, 147.4 mixture of 2 diastereomers; ^1H NMR (300MHz C_6D_6) δ 8.05 (m, 1H, thiazole H4), 7.90 (m, 1H, thiazolyl H5), 7.65 (m , 2H, H3, H5), 7.56-7.47 (m. 6H, Ar), 7.04-6.99 (m, 2H, Ar), 6.75-6.72 (m, 4H, Ar), 6.51 (m, 1H, H1'), 5.78 (m, 1H, H3'), 4.92 (m, 1H, H4'), 4.43 (m, 2H, H5'), 3.51-3.4 (m, 4H, CH_3-**CH** and O-**CH_2**,), 3.31 (m, 6H, Ar-O**CH_3**), 3.27-3.19 (m, 4H,), 2.98-2.86 (m, 1H, H2').2.41-2.19 (m, 1H, H2'), 1.79-1.61 (m) ,1.36-1.24 (m), 1.14-1.01 (m), 1.01-0.9(m) mixture of 2 diastereomer.

2.5 References

1) Johnson, W. T.; Zhang, P.; Bergstrom, D. E. *Nucleic Acids Res.* **1997**, *25*, 559-567.

2) Robins, M. J.; Vinayak, R. S.; Wood, S. G. *Tetrahedron Lett.* **1990**, *31*, 3731-3734.

3) Guiterrez, A. J.; Terhorst, T. J.; Matteucci, M. D.; Froehler, B. C. *J. Am. Chem. Soc.* **1994**, *116*, 5540-5544.

4) Peters, D.; Hornfeldt, A. B.; Gronowitz, S. *J. Heterocylic Chem.* **1990**, *27*, 2165-2173.

5) Huttle, R.; Buchele, F. *Chem. Ber.* **1955**, *88*, 1586-1590.

6) Caruthers, M. H.; Barone, A. D.; Beaucage, S. L.; Dodds, D. R.; Fisher, E. F.; McBride, L. J.; Matteucci, M.; Stabinsky, Z. *Meth. Enzymol.* **1985**, 287-313.

7) Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981**, *22*, 1859-1862.

8) McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett.* **1983**, *24*, 245-248.

9) Adams, S. P.; Kavka, K. S.; Wykes, E. J.; Holder, S. B.; Gallupo, G. R. *J. Am. Chem. Soc.* **1983**, *105*, 661-663.

10) Dahl, B. H.; Nielson, J.; Dahl, O. *Nucleic Acids Res.* **1987**, *4*, 1729-1743.

11) Bergstrom, D. E., personal communication of 5-nitroindole nucleoside extinction coefficient.

Chapter 3
Thermodynamic Stability of Oligonucleotides containing Pyrazole Nucleobase Analogs

3.1 Structure of DNA

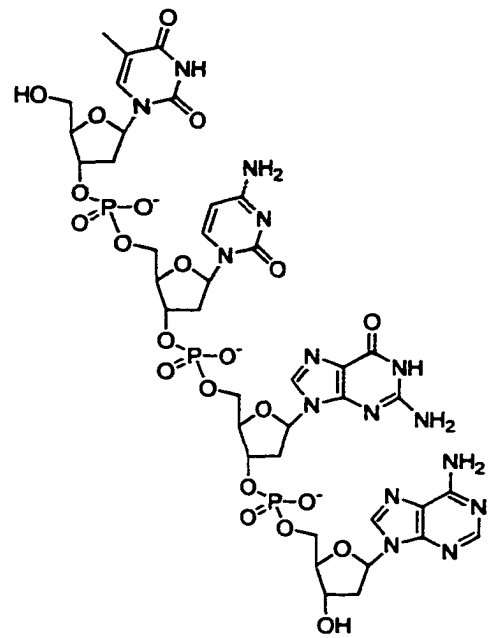


Figure 3.1 Structure of 5′-TCGA-3′ nucleotide

DNA is a polymer of nucleotides connected from the 3′ end of one base to the 5′ end of the next base (Figure 3.1). In nature DNA structure can be found as a double helix as discovered by Watson and Crick[1]. The double helix is comprised of two antiparallel DNA strands. The bases of one strand hydrogen bond with the bases on the opposite strand in the center of the helix. The bases are planar with the helical axis perpendicular to the base planes. The bases form hydrogen bonds across the plane specifically between purines and pyrimidines. Hydrogen bonds form "exclusively" between A and T and between G and C (Figure 3.2). Any other pairing between the two strands is considered a mismatch. Another interaction

between the bases is called base stacking. Base stacking occurs between bases perpendicular to the helix axis.(See Section 2.1) The aromatic systems of the bases offer π rich clouds as well as charge variation due to the positive and negative centers from the nitrogens and oxygens in the bases. (See Figure 2.2)

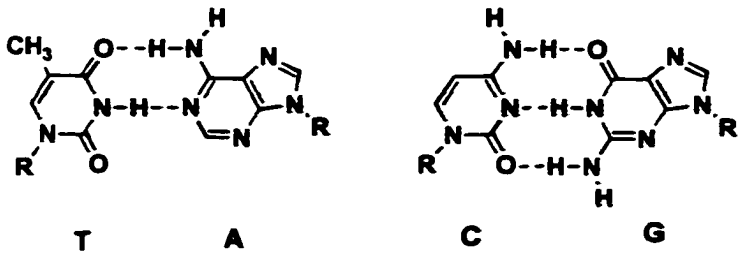


Figure 3.2 Hydrogen bond pairs between purines and pyrimidines

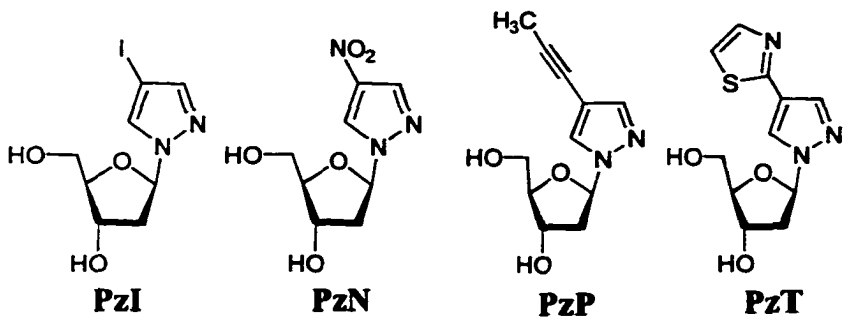


Figure 3.3. Structures of 4-substituted pyrazole nucleoside analogs

3.2 DNA Thermal Denaturation and Thermodynamic Analysis

3.2.1 Sequences Investigated

Two series of sequences were investigated. The first set of sequences were designed based on the self-complementary Dickerson sequence. The Dickerson sequence is a well known and studied sequences with many published NMR and crystal structures.

Table 3.1 Sequences for self-complementary sequences

X-A	5' CGC **Q**AA TT**A** GCG 3'
X-C	5' CGC **Q**AA TT**C** GCG 3'
X-G	5' CGC **Q**AA TT**G** GCG 3'
X-T	5' CGC **Q**AA TT**T** GCG 3'

Table 3.2 Sequences for non-self-complementary sequences

Sequence 1	5'CAA A**Q**T GGT GGC CAA GT 3' 3'GTT TTA CCA CCG GTT CA 5'
Sequence 2	5' CAA AAT GG**Q** GGC CAA GT 3' 3' GTT TTA CCA CCG GTT CA 5'
Sequence 3	5' CAA AA**Q** GG**Q** GGC **Q**AA GT 3' 3' GTT TTA CCA CCG GTT CA 5'
Sequence 4	5' CAA **QQ**T GGT GGC C**Q**A GT 3' 3' GTT TTA CCA CCG GTT CA 5'

3.2.2 Thermal Denaturation

Thermal denaturation of a DNA duplex is when the duplex is converted to two individual strands with heat. The point at which half of the helix is unwound our half of the duplex is in single stranded state is called the melting temperature or T_m. The T_m can be measured in several different ways. The method most often used is temperature-dependence UV absorbance spectroscopy and the second method is differential scanning calorimetry (DSC).[2] DSC monitors the change in

heat capacity while changing the temperature. The first method monitors the change in UV absorbance while changing the temperature. Each method analyzes the shape of or the area under the curve to provide the initial thermodynamic data.

3.2.3 Analysis of Data from Thermal Denaturation Studies

UV melting was performed as previously described.[3] The melting temperature, T_m, is defined as the point at which $\alpha = 0.5$, where α is defined as the mole fraction of single stranded DNA in the duplex state. It is assumed melting behavior is two state and there are no intermediates. In a non two-state transition, the melting curve will not have the characteristic shapeindicative of two state transtions. The equilibrium constant, at T_m, is given by the following,

$K = \frac{\alpha}{2(1-\alpha)^2 C_t}$ for self-complementary sequences and $K = \frac{2\alpha}{(1-\alpha)^2 C_t}$ for non-self complementary sequences. Analysis of melting data was performed with the program, **Meltwin©**, a non-linear least squares fitting program which determines T_m and thermodynamic data from single melting curves. For each sequence a series of melts were performed over at least 100-fold concentration range. To determine the thermodynamic parameters, ΔH_{VH} and ΔS_{VH}, van't Hoff style analysis was performed according to the following equations:

$$\frac{1}{T_m} = \left(\frac{R}{\Delta H}\right) \bullet \ln(C_t) + \frac{\Delta S}{\Delta H} \quad \text{for self-complementary sequences}$$

$$\frac{1}{T_m} = \left(\frac{R}{\Delta H}\right) \bullet \ln\left(\frac{C_t}{4}\right) + \frac{\Delta S}{\Delta H} \quad \text{for non-self complementary sequences}$$

The free energy, ΔG of the system was calculated at 298K using Gibb's equation:

$\Delta G = \Delta H - T\Delta S$.

3.3 Results

3.3.1 Self-Complementary Sequences

Table 3.3 Thermodynamic data for self-complementary sequences

X-Y	T_m °C (15 μM)	ΔT_m per mod. °C	Δ H (kcal/mol)	Δ S (cal/K•mol)	Δ G (kcal/mol at 298 K)
Con T-A	55.8		-69.5	-189.3	-13.1
Con G-C	67.0		-87.7	-235.9	-17.4
Con C-G	62.2		-69.1	-183.9	-14.3
Con A-T	61.9		-80.9	-219.3	-15.5
Ind-A	26.3	-14.8	-28.5	-73.1	-6.7
Ind –C	32.0	-17.5	-20.4	-44.8	-7.1
Ind –G	34.1	-14.1	-32.9	-84.9	-7.6
Ind –T	38.4	-11.7	-30.3	-75.3	-7.9
PzI-A	22.5	-16.7	-39.0	-110.0	-6.3
PzI-C	4.7	-31.1	-28.8	-81.7	-4.5
PzI-G	9.5	-26.3	-26.8	-72.8	-5.1
PzI –T	13.7	-24.1	-26.5	-70.4	-5.5
PzN-A	27.2	-14.3	-31.9	-84.3	-6.8
PzN-C	20.0	-23.5	-28.9	-76.6	-6.1
PzN-G	18.6	-21.8	-41.6	-120.4	-5.7
PzN-T	19.4	-21.2	-31.6	-85.8	-6.0
PzP-A	26.3	-14.7	-35.8	-97.6	-6.8

Table 3.3 continued

X-Y	T_m °C (15 μM)	ΔT_m per mod. °C	Δ H (kcal/mol)	Δ S (cal/K•mol)	Δ G (kcal/mol at 298 K)
PzP-C	15.4	-25.8	-41.3	-121.2	-5.2
PzP-G	26.0	-18.1	-43.0	-121.8	-6.7
PzP-T	15.7	-23.1	-25.3	-65.6	-5.8
PzT-A	33.4	-11.2	-47.6	-133.2	-7.9
PzT-C	24.0	-21.5	-32.7	-88.0	-6.5
PzT-G	36.2	-13.0	-42.9	-116.7	-8.2
PzT-T	27.9	-17.0	-33.8	-90.3	-6.9

3.3.2 Non-Self-Complementary Sequences

Table 3.4 Melting temperature (T_m) and thermodynamic parameters for non-self-complementary duplexes

Nucleoside-Sequence	T_m °C (15 μM)	ΔT_m per mod. °C	Δ H (kcal/mol)	Δ S (cal/K•mol)	Δ G (kcal/mol at 298 K)
Control	72.2		-172.5	-475.0	-31.0
Ind-1	71.6	-0.6	-106.4	-283.8	-21.8
Ind-2	65.9	-6.3	-125.2	-344.5	-22.5
Ind-3	48.2	-8.0	-108.9	-314.1	-15.2
Ind-4	51.9	-6.8	-52.4	-136.6	-11.7
PzN-1	66.2	-6.1	-66.1	-170.1	-15.4
PzN-2	64.8	-7.4	-119.0	-327.4	-21.4
PzN-3	41.0	-10.4	-78.2	-224.4	-11.4

Table 3.4 continued

X-Y	T_m °C (15 μM)	ΔT_m per mod. °C	Δ H (kcal/mol)	Δ S (cal/K•mol)	Δ G (kcal/mol at 298 K)
PzN-4	58.1	-4.7	-51.1	-129.5	-12.5
PzP-1	65.7	-6.6	-104.0	-282.1	-19.9
PzP-2	65.4	-6.9	-126.2	-348.2	-22.5
PzP-3	41.2	-10.3	-94.9	-277.3	-12.3
PzP-4	56.3	-5.3	-93.8	-259.9	-16.3
PzT-1	69.3	-2.9	-83.3	-218.4	-18.2
PzT-2	64.5	-7.8	-85.7	-229.2	-17.4
PzT-3	45.0	-9.1	-81.7	-232.2	-12.5
PzT-4	59.8	-4.2	-87.5	-238.2	-16.5

3.4 Discussion

3.4.1 Self-Complementary Sequences

We chose these Dickerson-based self-complementary sequences to be consistent with the analysis Bergstrom and coworkers have performed.[4] The general trend of our data agrees with their work[4] (Table 3.4) but our reported T_ms of sequences common between the two are generally higher. This is likely due to the difference in the methods used to analyze the data. They used the method of Carothers and Gralla[5] whereby $T_{1/2}$ and $T_{3/4}$ are used to determine ΔH from a single curve. In contrast, we have used the program **Meltwin©** which fits α throughout the curve via a *Marquardt-Levenberg* non-linear least squares algorithm to determine T_m. Additionally, we have used four Dickerson-like control sequences[4,6-8] for purposes of comparing a series of an analog across from all four

natural bases. This is particularly important in that the sequences with purines in the "Y" position, ConT-A and ConC-G (see Table 3.4), have T_ms ~5 degrees lower than the sequences in which the purines are in the "X" position (ConAT and ConGC resp.). Thus we analyze our data in terms of ΔT_m/substitution to take into account the different stacking energies already present in the parent sequences. As suggested by Freier,[9] we feel the ΔT_m analysis is probably more valid than comparison of thermodynamic parameters as the cooperativity of the melting transitions is variable with the different analogs.

The average ΔT_m's show the pyrazole analogs are much more stable when paired with purines than with pyrimidines. This is partly the case with **Ind** except with T. Possibly a variation in the syn or anti configuration of the indole ring could explain this phenomena with the lowest ΔT_m with T as its pairing partner. All the analogs show decreased stability increase with C as the pairing partner. In spite of Con-TA sequence having the lowest T_m of the four natural controls, each of the pyrazole analogs have the highest T_m when paired with A when compared to the other three complements. In each self-complementary data set, the pyrazole analog that destabilizes the duplex the least is **PzT**. This is probably due to the increased size and increased charge distribution the thiazole ring has over the other pyrazole substituents. Froehler has shown the propynyl unit substituted on pyrimidines[10] and purines[11] raises the T_m by approx. 1.5 °C and 1-1.4 °C per substitution, respectively. The average ΔT_m for the propynyl unit was -7°C, perhaps lack of the charge distribution or lower polarizability are responsible for the lower thermodynamic stability as well as the substituents being placed on a non-hydrogen

bonding base. **PzI** was extremely destabilizing as demonstrated by the very low T_m's. Thus, the iodo group provides no added stability over an unsubstitued pyrazole wich Bergstrom showed to be equivalent to an abasic site.[4] The nitro group is much less destabilizing than the iodo group. The smaller size and better charge distribution would make it a more suitable substituent. Comparing the T_m values for **PzN** to the values reported by Bergstrom,[4] the same trends are found with a slight variation in the actual numbers. This could be due to the different methods of analysis.

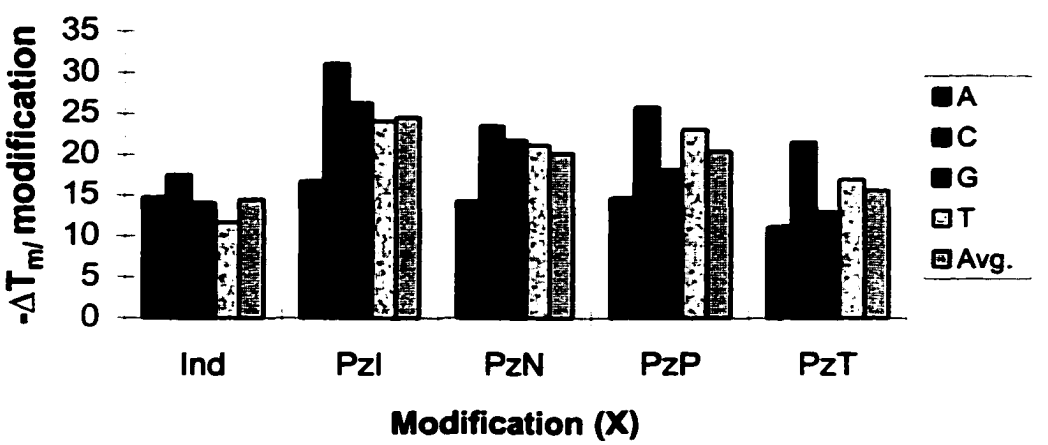


Figure 3.4 Comparison of $-\Delta T_m$ for each modification with all the natural bases

The pyrazole analogs overall are not as indiscriminate as Ind, but PzT is more stable than Ind when paired with A as well as with G. This may be related to the pairing partner size. It appears part of the reason for the pairing preferences may be an attempt at maintaining the traditional size of the natural pyrimidine-purine pair. The pyrazole analogs prefer pairing with A. When the pyrazole analogs are paired with A, they are replacing a pyrimidine, comparable in size, in center of the duplex.

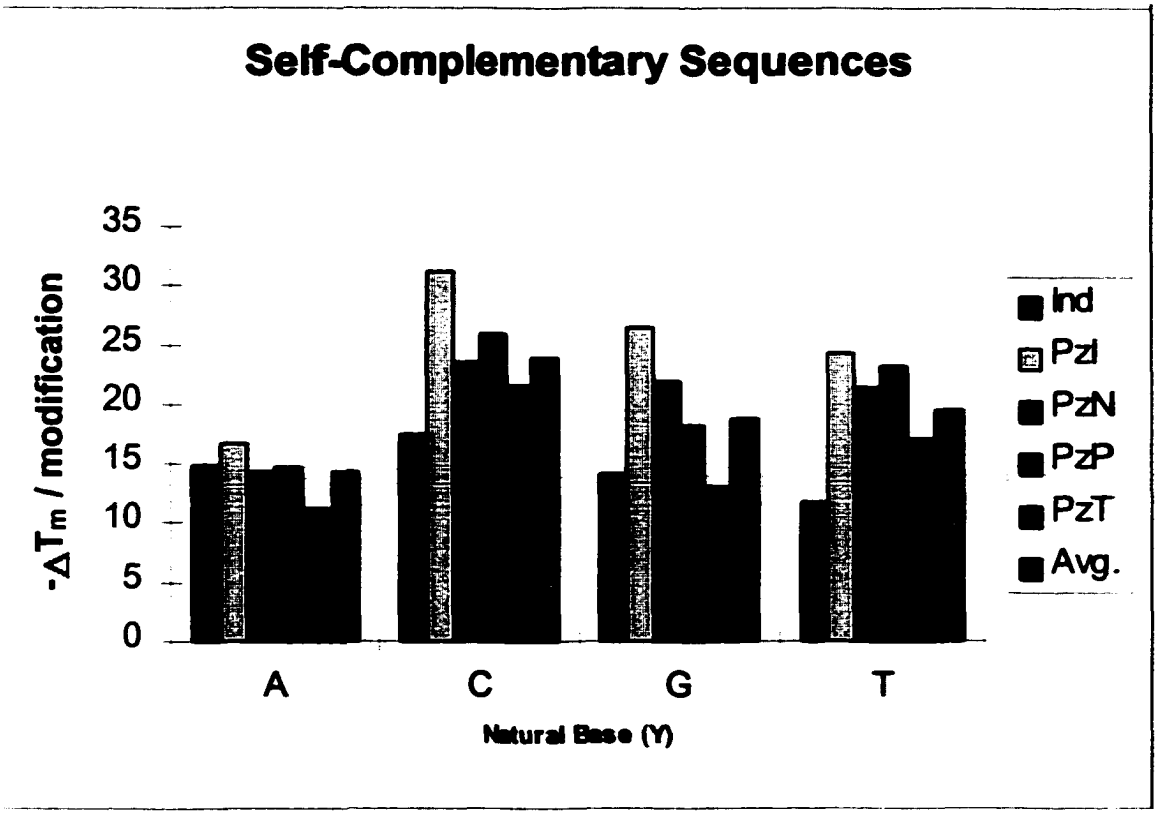


Figure 3.5 Comparison of ΔT_m for each natural base with all of the modifications

This maintains the traditional purine-pyrimidine size in the duplex. This may also explain the increase in the ΔT_m. When the analogs are paired with C, the pyrazole analogs are replacing G in the sequence, offering a "pyrimidine-pyrimidine" or mismatch in size in the center of the duplex, therefore much more destabilizing to the duplex than the natural complement. The pyrazole analogs are less destabilizing to the duplex than Kool's analogs 4,6-dimethyindole and 2,4-diflourotoulene. Both of these analogs are very non-polar and do not offer the same charge variation the pyrazole analogs offer. 4,6-Dimethylindole, which is designed as an A analog, does not thermodynamically favor pairing with T and shows the following preference C> A> G> T.[12] 2,4-Difluorotoluene was found to favor

pairing with C while it was designed as a T mimic and shows the following thermodynamically favored pairing preferences C > G> A> T.[12] As both these analog are favoring pairing with C they are "replacing" a G in the natural sequence

3.4.2 Non-Self Complementary Sequences

Previously,[13-15] it has been shown a single modification in the center of a duplex is more destabilizing than a modification near an end. Our data (Table 3.5), Sequence 1 vs. Sequence 2 and Sequence 3 vs. Sequence 4, supports this observation as well. As expected from the self-complementary data, **PzT** modification appears to be the least destabilizing of the pyrazole modifications but not as stable as **Ind**. The results from Seq. 4 show that the pyrazole bases are able to stack side well side-by-side. Bergstrom has suggested a ring nitrogen and its ability to hydrogen bond if in the correct position makes the base more discriminating. The imidazole N3 when paired with G, according to models, is in a position capable of H-bonding similar to a C without disturbance of the duplex,as suggested by Bergstrom et al. The ring nitrogen for pyrazole is in the 2 position and would require some distortion of the duplex to H-bond with G and additionall it is a much weaker hydrogen-bond acceptor than an imidazole N-3. The natural bases have charge centers in different positions abling them to stack with favorable charge interaction with the bases above and below. Some of the analogs studied although π-rich, lack the charge distribution and are not good stacking bases. The thiazole analog provides both positive and negative centers which can improve its ability to stack, as well as the π-systems of the two 5 membered rings, which may explain higher T_ms of **PzT** containing oligonucleotides.

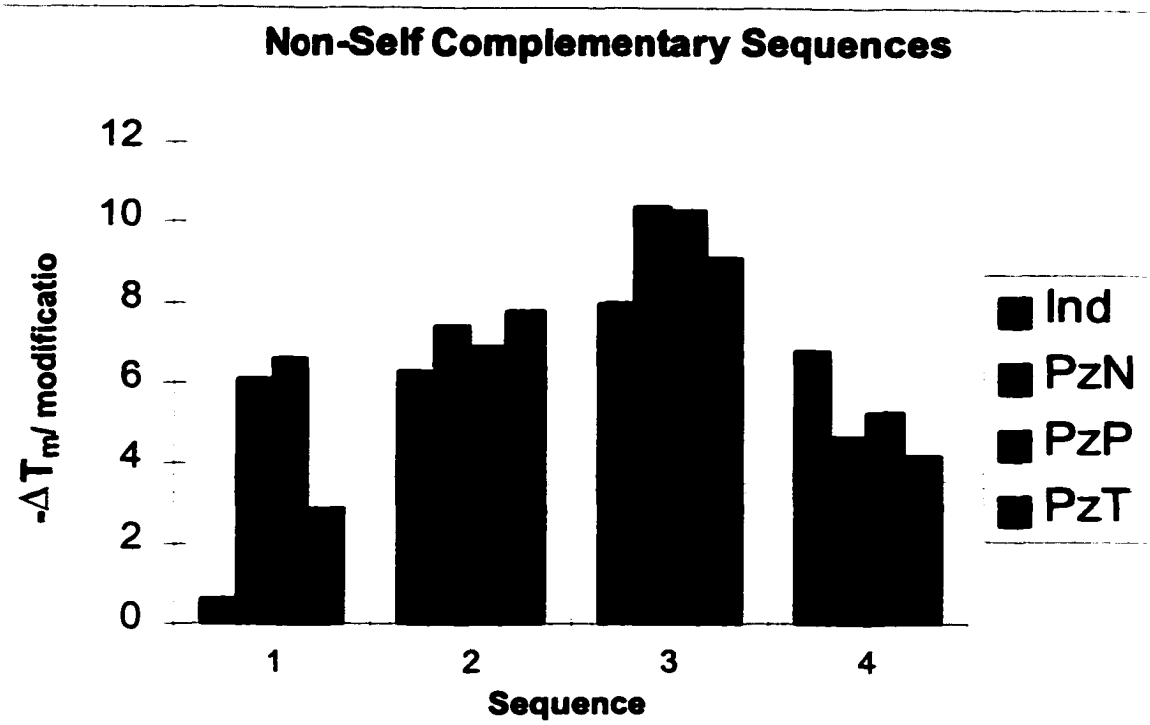


Figure 3.6 Comparison of ΔT_m for non-self-complementary sequences

A possible explanation for the very low $-\Delta$Tm for **Ind** in Sequence 1 when compared to the pyrazole analogs could be due to preferences in pairing partners. The self-complementary data shows **Ind** prefers pairing with pyrimidines over purines. The self-complementary data also shows the pyrazole analogs do not favor pairing with pyrimidines. Sequence 1 has the modification across from C. As **Ind** favors this pairing and the pyrazole analogs do not this might explain the large difference in the $-\Delta T_m$ shown in Figure 3.6.

3.5 Summary

Comparison of the data for both each sequences show the pyrazole analogs have pairing preference for the natural base A. The analogs are capable of being incorporated into a duplex without disrupting the natural order of the duplex. The thiazolyl substituent shows the best ability to stabilize the duplex when compared to

the other substituents. The pyrazole analogs are not a indiscriminate as **Ind** when pairing with the four natural bases.

3.6 Experimental

3.6.1 UV Thermal Denaturing Measurements.

Absorbance measurments were recorded at 260 nm on a Gilford Response UV/VIS Spectrophotomer equipped with a Peltier temperature controlling device and thermal software while the temperature was raised from 5 °C to 95 °C in 0.5 °C increments under an inert atmosphere. To determine that evaporation was less than 5%, the average of 3-5 absorbance readings was determined at 5 °C and/or 25 °C before and after each melt. Prior to measurement all samples were heated to 95°C for 2 min to drive the equilibrium to single strand. They were then cooled to 5°C and equilibrated for 5 min to allow for duplex formation. One concentration for each duplex was determined in both a heating and cooling ramp to ensure that the duplex behaved the same way. Then 3-8 concentrations (covering at least a 100-fold concentration difference) were run in a heating run at 1 °C / min.

3.6.2 Solvent Preparation

An estimate of the amount of oligonucleotide needed was determined by UV(260 nm) (Gilford) in H_2O and the sample was evaporated to dryness. The DNA duplexes were dissolved in water and had a final buffer concentration of 1 M NaCl, 1 mM EDTA, 10 mM Na_2HPO_4. The cell pathlength varied from 0.01 cm to 1 cm. All solvents were degassed in a vacuum desiccator prior to use. The concentration was determined using the assumption at high temperature the oligonucleotides are unstacked and unpaired. The absorbance was determined at 85

°C and Beer's Law was used to determine the concentration utilizing an estimate of the extinction coefficients as calculated below.

3.6.3 Extinction Coefficients for Oligonucleotides

The nearest neighbor estimate for each non-modified sequences was determined using the method and values from the literature.[16,17] The straight sum of the extinction of the mononucleotides in each non-modified sequence was made. The ratio of nearest neighbor values: straight sum method was calculated. The ratio was then multiplied by the straight sum of the individual mononucleotides (incorporating the determined extinction coefficients of the modified nucleosides) for the modified sequences. This will provide a nearest neighbor estimate for the extinction coefficient as the extinction coefficient for each dinucleotide needed for nearest neighbor estimates is not known.

3.6.4 Calculation of Thermodynamic Parameters

The T_m as well as ΔH, ΔS and ΔG were determined for each concentration using **Meltwin©** provided by Jeffrey McDowell.[18] This program fits a curve to the raw data points. The parameters from the curve allow for determination of the T_m, ΔH, ΔS and calculation of ΔG for each individual concentration. Once analysis was completed for three or more concentrations, van't Hoff analysis was done to provide the thermodynamic values reported. The program provided T_m at a concentration of 100 μM and ΔG at 310 K, using the $1/T_m$ and $\ln C_t/C_o$ relationship for self and non-self complementary sequences the T_m's were converted to be reported at 15 μM and ΔG at 298 K using the relationship of $\Delta G = \Delta H - T\Delta S$ and $\Delta G = -RT \ln K$,

rearrangement results in the following equation $K = \exp\left(\frac{\Delta S}{R} - \frac{\Delta H}{RT}\right)$, which allows for van't Hoff analysis of the linear relationship between $1/T_m$ and ln C_t/C_o to be used.

3.6.5 DNA Melting Experiments and Thermodynamic Analysis

UV melting was performed as previously described.[3] The melting temperature, T_m, is defined as the point at which α =0.5 when α is defined as the mole fraction of single stranded DNA in the duplex state. It is assumed melting behavior is two state and there are no intermediates. The equilibrium constant, at T_m, is given by the following, $K = \frac{\alpha}{2(1-\alpha)^2 C_t}$ for self complementary sequences and $K = \frac{2\alpha}{(1-\alpha)^2 C_t}$ for non-self complementary sequences. Analysis of melting data was performed with the program, **Meltwin©**, a non-linear least squares fitting program which determines T_m and thermodynamic data from single melting curves. For each sequence a series of melts were performed over at least 100-fold concentration range. To determine the thermodynamic parameters, ΔH_{VH} and ΔS_{VH}, van't Hoff style analysis was performed according to the following equations:

$$\frac{1}{T_m} = \left(\frac{R}{\Delta H}\right) \bullet \ln\left(\frac{C_t}{C_o}\right) + \frac{\Delta S}{\Delta H} \quad \text{for self-complementary sequences}$$

$$\frac{1}{T_m} = \left(\frac{R}{\Delta H}\right) \bullet \ln\left(\frac{\frac{C_t}{C_o}}{4}\right) + \frac{\Delta S}{\Delta H} \quad \text{for non-self complementary sequences}$$

The free energy, ΔG, was calculated from the determined ΔH_{VH} and ΔS_{VH} using the equation: $\Delta G = \Delta H - T\Delta S$.

Table 3.5 Self-Complementary Sequence Data for each analysis

Sequence	Conc. μM	ΔH kcal/mol	ΔS cal/mol K	ΔG kcal/mol	T_m °C	R value
Cont TA	3.6	-50.15	-129.77	-9.91	51.06	0.998
	12.4	-57.5	-152.58	-10.18	55.66	
	90.4	-76.26	-208.92	-11.47	62.19	
	268.0	-78.03	-214.47	-11.51	64.92	
	599.0	-78.63	-216.3	-11.54	67.17	
Con GC	9.3	-70.91	-186.45	-13.08	65.35	0.997
	17.9	-75.1	-198.86	-13.43	67.33	
	98.9	-83.62	-223.66	-14.26	72.43	
	9	-76.02	-201.3	-13.58	65.62	
	41.9	-82.7	-220.92	-14.18	70.09	
	110.0	-82.94	-221.84	-14.14	72.5	
	922.0	-99.88	-270.81	-15.89	77.69	
Cont C-G	15.7	-46.06	-115.37	-10.27	62.15	0.999
	32.8	-54.89	-141.83	-10.9	64.97	
	118.0	-62.74	-165.3	-11.47	69.2	
	764.0	-82.06	-221.21	-13.45	75.33	
Cont A-T	10.1	-61.99	-162.81	-11.49	60.7	0.999
	15.8	-65.73	-174.07	-11.74	62.12	
	635.0	-81.21	-220.53	-12.81	72.18	

Table 3.5 continued

	759.0	-74.14	-199.67	-12.21	73.38	
Ind-A	6.2	-35.66	-97.59	-5.39	20.56	0.989
	12.5	-30.8	-81.41	-5.55	23.43	
	47.2	-53.01	-151.27	-6.1	36.76	
	953.0	-42.29	-115.66	-6.42	53.51	
Ind-C	3.5	-24.33	-58.47	-6.19	18.5	0.992
	18.7	-27.29	-67.14	-6.47	34.3	
	38.1	-43.49	-117.24	-7.13	43.2	
	145.1	-52.58	-144.1	-7.91	52.2	
Ind-G	7.1	-30.23	-76.86	-6.39	27.88	0.99
	14.2	-30.17	-75.76	-6.68	34.94	
	50.4	-43.44	-117.8	-6.9	42.84	
	908.0	-98.48	-283.1	-10.68	58.43	
Ind-T	13.8	-32.59	-82.81	-6.9	37.06	0.986
	30.0	-37.6	-97.8	-7.26	44.1	
	103.0	-43.33	-115.6	-7.48	50.6	
PzI-A	4.9	-38.59	-108.77	-4.86	16.88	0.995
	11.3	-47.48	-138	-4.68	22.43	
	22.8	-42.55	-121.79	-4.78	24.36	
	82.7	-42.53	-121.73	-4.77	29.73	
	445.0	-50.71	-147.37	-5	38.52	
PzI-C	15.2	-23.12	-61.05	-4.18	5.02	0.999

Table 3.5 continued

	70.5	-24.89	-68.12	-3.76	12.52	
	423.0	-29.53	-83.87	-3.52	24.26	
	1155.0	-28.24	-79.8	-3.49	29.68	
PzI-G	8.4	-19.22	-45.6	-5.08	6.15	0.994
	30.1	-19.64	-47.12	-5.03	16.51	
	72.6	-24.99	-66.48	-4.37	19.37	
	814.5	-37.25	-106.28	-4.29	36.23	
PzI-T	13.0	-25.99	-68.85	-4.64	11.83	0.988
	19.5	-23.53	-59.79	-4.99	16.17	
	90.3	-31.14	-85.34	-4.67	26.73	
	210.5	-29.22	-79.61	-4.53	29.85	
PzN-A	5.9	-35.91	-97.94	-5.53	21.55	0.994
	9.6	-37.1	-101.15	-5.69	25.53	
	53.7	-45.53	-128.5	-5.67	34.39	
PzN-C	5.2	-37.84	-107.36	-4.55	14.55	0.996
	10.4	-26.25	-67.6	-5.28	17.04	
	48.9	-34.51	-95.13	-5.01	27.36	
	71.2	-44.77	-129.5	-4.61	28.39	
	184.0	-59.74	-175.9	-5.19	36.41	
PzN-G	7.9	-37.98	-108.44	-4.35	15.03	0.992
	15.2	-37.36	-105.36	-4.68	20.09	
	91.2	-40.9	-118.35	-4.2	25.79	

Table 3.5 continued

	701.0	-38.9	-109.1	-4.25	35.16	
PzN-T	6.8	-21.45	-50.93	-5.65	14.55	0.996
	19.9	-28.03	-73.39	-5.27	22.22	
	95.3	-33.58	-92.74	-4.81	28.97	
	805.0	-43.17	-122.4	-5.16	42.64	
PzP-A	7.6	-35.69	-96.86	-5.64	23.54	0.997
	13.8	-33.38	-89.63	-5.59	25.28	
	24.6	-37.43	-102.75	-5.56	29.06	
	194.2	-33.97	-91.36	-5.63	40.39	
	295.8	-49.35	-140.62	-5.74	41.66	
	578.1	-69.61	-203.43	-6.52	45.84	
PzP-C	12.6	-16.43	-34.7	-5.67	14.47	0.998
	80.2	-18.04	-42.19	-4.95	22.84	
	225.0	-19.66	-48.91	-4.49	26.51	
	430.0	-24.44	-65.4	-4.16	29.3	
PzP-G	9.6	-30.58	-80.19	-5.71	23.32	0.997
	17.4	-36.96	-101.1	-5.6	27.61	
	1566.0	-49.99	-143.62	-5.45	46.4	
PzP-T	8.5	-19.04	-43.5	-5.54	12.24	0.999
	59.2	-31.81	-87.62	-4.63	24.24	
	277.0	-47.28	-13.643	-4.97	36.47	
PzT-A	14.4	-39.82	-108.29	-6.24	32.13	0.989

Table 3.5 continued

	23.9	-42.89	-117.68	-6.4	35.82	
	91.2	-49.39	-138.65	-6.38	41.13	
	210.4	-50.58	-142.13	-6.5	45.08	
	775.1	-53.84	-153.09	-6.36	48.65	
PzT-C	3.9	-36.27	-101.3	-4.85	14.61	0.977
	11.4	-33.74	-90.23	-5.75	25.81	
	478.0	-88.75	-269.1	-5.29	36.45	
	314.0	-40.85	-114.23	-5.42	40.47	
PzT-G	1.3	-41.33	-111.82	-6.65	24.7	0.994
	7.3	-42.42	-114.65	-6.87	33.91	
	18.5	-47.89	-132.19	-6.89	38.15	
	56.5	-53.61	-150.19	-7.02	42.88	
	133.8	-44.83	-122.86	-6.72	45.72	
	245.7	-55.02	-154.63	-7.06	48.33	
PzT-T	4.0	-30.83	-80.12	-5.98	20.94	0.995
	14.6	-26.75	-66.74	-6.05	27.87	
	70.7	-32.56	-86.4	-5.76	35.76	
	176.0	-50.29	-142.26	-6.16	42.24	
	388.0	-54.75	-154.9	-6.71	47.98	

Table 3.6 Non-Self-Complementary Sequence Data for each analysis

Sequence	Conc. μM	ΔH kcal/mol	ΔS cal/mol K	ΔG kcal/mol	T_m °C	R value
Control	2.03	-160.06	-438.66	-24.01	69.24	0.998
	15.7	-117.53	-315.52	-19.67	72.25	
	26.9	-120.62	-324.72	-19.91	72.93	
	63.6	-112.95	-303.3	-18.89	74.13	
	174.8	-102.76	-273.94	-17.49	75.33	
Ind-1	0.42	-104.6	-278.48	-18.24	63.86	0.999
	5.29	-109.81	-294.16	-18.58	68.88	
	55.4	-96.19	-254.42	-17.28	74.55	
Ind-2	3.01	-119.73	-328.05	-17.99	63.12	0.997
	28.8	-97.6	-263.83	-15.77	66.5	
	533	-54.27	-139.21	-11.09	72.61	
Ind-3	0.156	-68.12	-184.19	-10.99	39.2	0.993
	1.18	-55.68	-145.99	-10.4	43.43	
	12.48	-53.48	-141.02	-9.74	48.59	
	216.9	-34.99	-87.93	-7.72	52.5	
Ind-4	0.351	-13.63	-11.56	-10.04	37.63	0.999
	4.49	-21.93	-41.39	-9.09	46.43	
	38.9	-27.97	-62.08	-8.72	55.93	
Ind-1	2.52	-74.43	-194.67	-14.05	60.55	0.991
	12.7	-66.05	-170.54	-13.16	64.37	

Table 3.6 continued

	120.7	-108.39	-291.7	-17.92	73.84	
PzN-2	0.562	-95.05	-255.53	-15.8	58.18	0.989
	2.5	-123.25	-339.6	-17.92	61.77	
	6.41	-100.15	-271.61	-15.91	62.77	
	11.4	-111.05	-303.41	-16.95	64.61	
	22	-108.83	-297.05	-16.7	65.77	
	54.6	-100.61	-273.84	-15.68	66.64	
PzN-3	1.54	-62.69	-174.09	-8.69	34.98	
	7.94	-63.52	-177.01	-8.62	39.58	
	75.3	-69.46	-196.91	-8.39	44.72	
PzN-4	0.895	-61.43	-160.81	-11.55	48.07	0.996
	7.6	-55.99	-144.11	-11.29	55.65	
	12	-61.8	-160.71	-11.96	59.15	
PzP-1	0.519	-126.78	-350.42	-18.09	58.82	0.991
	10.6	-117.95	-315.22	-17.18	59.73	
	3.96	-91	-262	-15.74	61.93	
	36.4	-56.9	-143.78	-12.3	67.87	
PzP-2	0.813	-141.83	-394.83	-19.37	60.2	0.999
	1.75	-156.3	-438.31	-20.36	61.26	
	8.045	-156.93	-439.55	-20.6	63.89	
	77.7	-141.75	-393.58	-19.68	68.3	
PzP-3	0.95	-73.27	-207.19	-9.01	35.35	0.999

Table 3.6 continued

	1.79	-85.61	-247.18	-8.94	36.76	
	8.63	-91.52	-266.09	-8.99	40.24	
	81.6	-81.12	-233.91	-8.57	44.49	
PzP-4	0.99	-83.68	-229.1	-12.63	49.56	0.998
	2	-73.46	-196.97	-12.37	52.2	
	891	-66	-178.02	-10.79	65.75	
PzT-1	0.252	-90.04	-238.26	-16.15	58.87	0.988
	1.94	-103.97	-280.12	-17.09	63.31	
	10.5	-99.88	-267.57	-16.89	67.61	
	17.9	-102.3	-275.04	-17	68.42	
PzT-2	0.3	-64.79	-165.82	-13.36	53.39	0.981
	2.67	-106.22	-289.26	-16.5	61.38	
	11.1	-103.64	-282.14	-1614	63.83	
	11.2	-104.33	-284.36	-16.13	63.64	
	20.92	-57.11	-144.98	-12.14	64.46	
	42.86	-102.36	-278.53	-15.97	66.61	
PzT-3	0.787	-63.46	-173.64	-9.6	37.42	0.991
	2.65	-71.3	-198.99	-9.59	40.6	
	7.09	-74.33	-208.13	-9.78	43.88	
	12.24	-70.83	-197.57	-9.55	44.75	
	23.84	-68.64	-191.57	-9.22	45.4	
	69.5	-77.28	-218.47	-9.52	48.52	

Table3.6 continued

PzT-4	0.71	-69.13	-182.47	-12.54	50.88	0.992
	2.81	-71.74	-190.27	-12.72	55.28	
	12.9	-55.36	-142.03	-11.31	58.07	
	25.2	-66.99	-177.31	-11.99	59.94	

3.7 References

1) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737-738.

2) Marky, L. A.; Breslauer, K. J. *Biopolymers* **1987**, *26*, 1601-1620.

3) LeBlanc, D. A.; Morden, K. M. *Biochemsitry* **1991**, *30*, 4042-4047.

4) Bergstrom, D. E.; Zhang, P.; Johnson, W. T. *Nucleic Acids Res.* **1997**, *25*, 1935-1942.

5) Caruthers, M. H.; Barone, A. D.; Beaucage, S. L.; Dodds, D. R.; Fisher, E. F.; McBride, L. J.; Matteucci, M.; Stabinsky, Z. *Meth. Enzymol.* **1985**, 287-313.

6) Ahmadian, M. H.; Zhang, P.; Bergstrom, D. E. *Nucleic Acids Res.* **1998**, *26*, 3127-3135.

7) Johnson, W. T.; Zhang, P.; Bergstrom, D. E. *Nucleic Acids Res.* **1997**, *25*, 559-567.

8) Zhang, P.; Johnson, W. T.; Klewer, D.; Paul, N.; Hoops, G.; Davisson, V. J.; Bergstrom, D. E. *Nucleic Acids Res.* **1998**, *26*, 2208-2215.

9) Freier, S. M.; Altman, K. H. *Nucleic Acids Res.* **1997**, *25*, 4429-4443.

10) Froehler, B. C.; Wadwani, S.; Terhorst, T. J.; Gerrard, S. R. *Tetrahedron Lett.* **1992**, *33*, 5307-5310.

11) Buhr, C. A.; Wagner, R. W.; Grant, D.; Froehler, B. C. *Nucleic Acids Res.* **1996**, *24*, 2974-2980.

12) Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863-1872.

13) Martin, F. H.; Castro, M. M.; Aboul-ela, F.; Ohtsuka, E. *Nucleic Acids Res.* **1986**, *14*, 7727-7736.

14) Kaimya, H.; Sakaguchi, T.; Murata, N.; Fujimuro, M.; Miura, H.; Ishikawa, K.; Shimizu, M.; Inoue, H.; Nishimura, S.; Matsukage, A.; Masutani, C.; Hanaoka, F.; Ohtsuka, E. *Chem. Pharm. Bull.* **1992**, *40*, 2792-2795.

15) Bischofberger, N.; Matteucci, M. D. *J. Am. Chem. Soc.* **1989**, *111*, 3041-3046.

16) Cantor, C. R.; Warshaw, M. M.; Shapiro, H. *Biopolymers* **1970**, *9*, 1059.

17) Warshaw, M. M.; Tinoco Jr., I. *J. Mol. Biol.* **1966**, *26*, 29.

18) McDowell, J. A.; Turner, D. *Biochemistry* **1996**, *35*, 14077-14089.

Chapter 4
Enzymic Behavior of Pyrazole-Containing Nucleobase Analogs with Various Thermostable DNA Polymerases

4.1 Target for PCR Reactions

The goal of this research is to determine the ability of polymerases to recognize the pyrazole analogs. We studied 4-nitropyrazole (**PzN**) and 4-(2-thiazolyl)-pyrazole (**PzT**)of the pyrazole analogs as well as 2 other well-studied modifications, 5-nitroindole (**Ind**) and 3-nitropyrrole (**PyN**). (Figure 4.1) Two different control experiments were designed. The first control experiment uses a non-modified primer containing the natural complement (G) to the position on the template. The second control experiment uses 4 identical primers except in the position of interest. Each primer contains one of each of the four natural nucleosides. These four primers are added in equal mole amounts to allow for quantitation of a mixture that should be 25% of each.

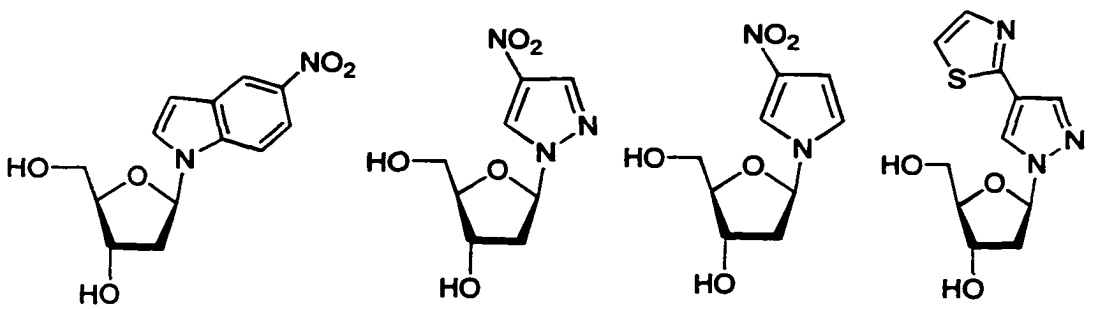


Figure 4.1 Analogs studied in PCR reactions

4.2 Design of Enzymic Experiment

Using a procedure originally described at Purdue University by Davisson and coworkers,[1] the expression vector *phisF-tac* plasmid[2] is digested using the

restriction endonuclease, Pvu II. One of the fragments from this digestion contains the *hisF* gene indicated in capital letters (Figure 4.3). This gene serves as part of a 797 bp template for the PCR amplification experiments. The *Pvu* II recognition site is 5'-CAG^CTG-3'. When this sequence is present the enzyme will cut the dsDNA between the G and C. The sense primer (40 bases) is designed for positions -10 to 30 of the *hisF* gene (underline on the top portion of Figure 4.3). The modification is in position 26. The antisense primer (30 bases) is corresponds to positions 757 to +10 (underlined on the bottom of Figure 4.3). The *hisF* gene is shown in capital letters. The modifications were incorporated using standard automated DNA synthesis once the phosphoramidites of the analogs were either purchased or synthesized as described in Chapter 2. During PCR the modified primers are initially extended resulting in full-length template containing the modification. The next round of PCR actually incorporates a natural nucleoside across from the pyrazole analog. Figure 4.2 is a pictorial representation of this process. This figure is very similar to Figure 1.7 describing PCR except for the pyrazole analogs present in the sense primers. Several different enzymes were chosen for this study. These include Ampli*Taq*® DNA polymerase (Perkin-Elmer), Vent DNA polymerase (New England Biolabs), Deep Vent DNA polymerase (New England Biolabs), Pfu DNA polymerase exo+, *Pfu* DNA polymerase exo-, and *UlTma* DNA polymerase (Perkin Elmer). As enzymes have different inherent abilities a wide variety of enzymes were chosen to determine the directing ability of the analogs. The question to be answered was whether the analogs directed specifically for one base or was it dependent on the enzyme's inherent properties. See Table 4.1 for a list of the exonuclease properties

of the DNA polymerases studied. The different type of exonuclease activities is discussed in Section 1.5.4.

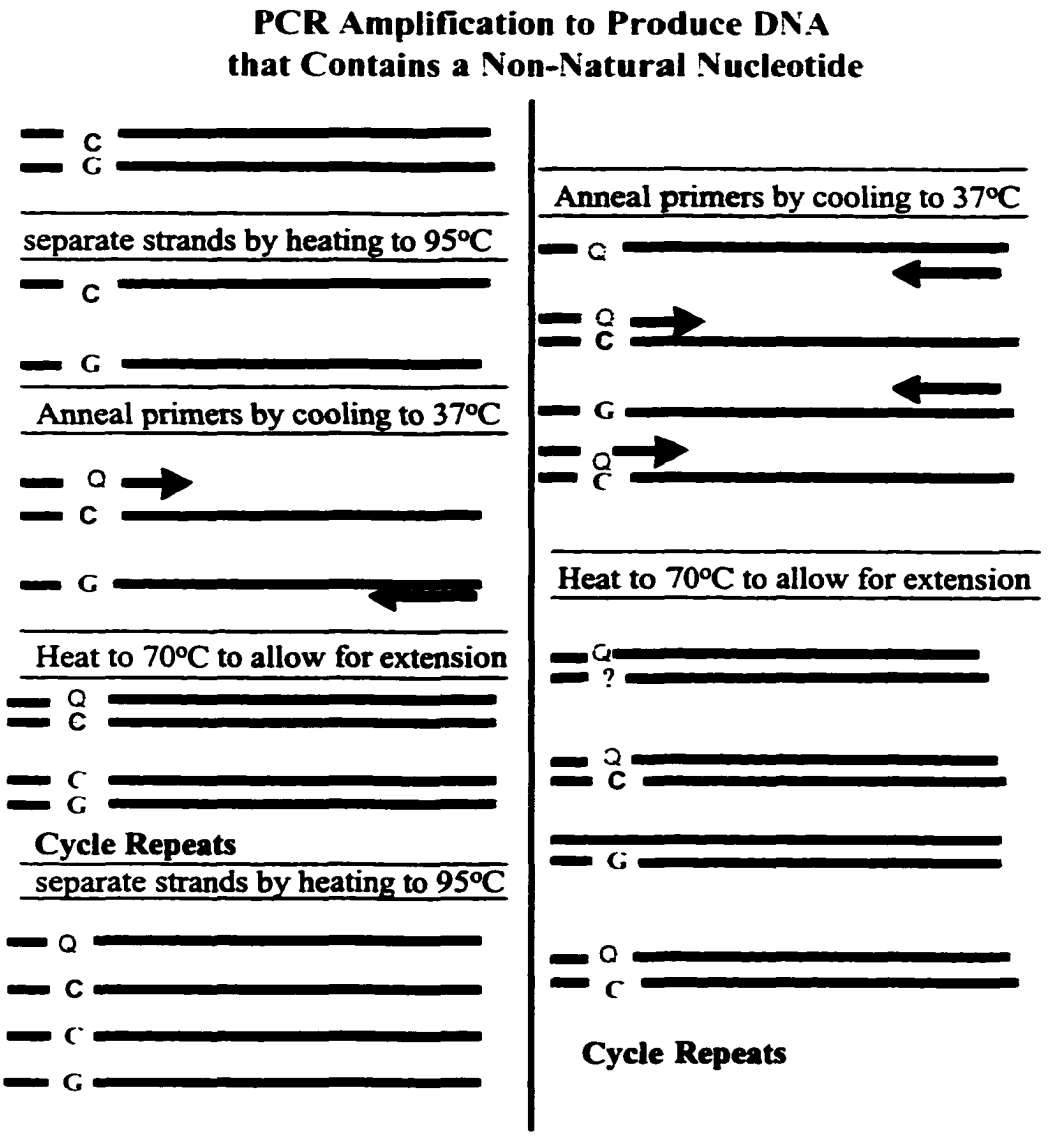


Figure 4.2 Diagram of PCR reaction containing a pyrazole analog in the sense primer

The fidelity of these enzymes is described by the number of mutations/ bp/ duplication.[3] *UlTma* is a very poor proofreading enzyme resulting in such a fidelity it is almost considered a non-proofreading enzyme. The best proofreading enzyme studied was *Pfu* exo+ followed by Vent and then Deep Vent.

aggaattcatATGCTGGCAAAACGCATAATCCCATGTCTCGACGTTCGTGATGGTCAGGT
tccttaagtaTACGACCGTTTTGCGTATTAGGGTACAGAGCTGCAAGCACTACCAGTCCA

CCACTTTCCCGTACAGTTTCGCAACCATGAAATCATTGGCGATATCGTGCCGCTGGCAAAA
GGTGAAAGGGCATGTCAAAGCGTTGGTACTTTAGTAACCGCTATAGCACGGCGACCGTTTT

CGCTACGCTGCGATGCGAGCTGACGAACTGGTGTTCTACGATATCACCGCTTCCAGCGATG
GAAGAAGGCCTTCTTCCGCGACTGCTTGACCACAAGATGCTATAGTGGCGAAGGTCGCTAC

GCCGTGTGGTAGATAAAAGCTGGGTATCTCGCGTGGCGGAAGTGATCGACATTCCGTTTTG
CGGCACACCATCTATTTTCGACCCATAGAGCGCACCGCCTTCACTAGCTGTAAGGCAAAAC

TGTGGCGGGTGGGATTAAGTCTCTGGAAGATGCCGCGAAAATTCTTTCCTTTGGCGCGGAT
ACACCGCCCACCCTAATTCAGAGACCTTCTACGGCGCTTTTAAGAAAGGAAACCGCGCCTA

AAAATTTCCATCAACTCTCCTGCGCTGGCAGACCCAACATTAATTACTCGCCTGGCCGATC
TTTTAAAGGTAGTTGAGAGGACGCGACCGTCTGGGTTGTAATTAATGAGCGGACCGGCTAG

GCTTTGGCGTGCAGTGTATTGTGGTCGGTATTGATGGAAACGCTCGACTGGGTACAGGAAG
CGAAACCGCACGTCACATAACACCAGCCATAACTACCTTTGCGAGCTGACCCATGTCCTTC

TGCAAAAACGCGGTGCCGGAGAAATCGTCCTCAATATGATGAATCAGGACGGCGTGCGTAA
ACGTTTTTGCGCCACGGCCTCTTTAGCAGGAGTTATACTACTTAGTCCTGCCGCACGCATT

CGGTTACGACCTCGAACAACTGAAAAAAGTGCGTGAAGTTTGCCACGTCCCGCTGATTGCC
GCCAATGCTGGAGCTTGTTGACTTTTTTCACGCACTTCAAACGGTGCAGGGCGACTAACGG

TCCGGTGGCGCGGGCACCATGGAACACTTCCTCGAAGCCTTCCGCGATGCCGACGTTGACG
AGGCCACCGCGCCCGTGGTACCTTGTGAAGGAGCTTCGGAAGGCGCTACGGCTGCAACTGC

GCGCGCTGGCAGCTTCCGTATTCCACAAACAAATAATCAATATTGGTGAATTAAAAGCGTA
CGCGCGACCGTCGAAGGCATAAGGTGTTTGTTTATTAGTTATAACCACTTAATTTTCGCAT

CCTGGCAACACAGGGCGTGGAGATCAGGATATGTTAacccgggtacg(797)
GGACCGTTGTGTCCCGCACCTCTAGTCCTATACAATtgggcccatgc

Figure 4.3 Sequence of the target PCR product, the *hisF* gene.[1]

Table 4.1 Exonuclease properties of DNA polymerases

DNA Polymerase	3′- 5′ exonuclease (proof-reading)	5′- 3′ exonuclease	Fidelity[3]
AmpliTaq®	-	+	8×10^{-06}
Pfu exo -	-	+	5×10^{-05}
UlTma®	+	+	5×10^{-05}
Pfu exo+	+	+	1.3×10^{-06}
Vent	+	+	2.7×10^{-06}
Deep Vent	+	+	2.8×10^{-06}

4.3 Analysis of PCR Products

The technique most often used to analyze the completed PCR reaction is gel electrophoresis using a sieving medium such as agarose. The crude PCR reaction will have unused primers (30mer and 40mer) and full length product (797bp) as well as buffers present. A small portion of the PCR reaction is combined with a loading buffer consisting of 2 dyes which migrate at different rates and a reagent Ficoll which is used to "weigh" down the solution to aid in loading the samples into the wells. Electrophoresis is used to separate the DNA to determine the size of the DNA present in the PCR reaction. Size markers are used to determine the size of the DNA found in the PCR reaction. Ethidium bromide is used to visualize the agarose gel. Ethidium bromide is a DNA intercalator and is visible under UV light when intercalated into DNA. An inverted image of an agarose gel is present in Figure 4.5, Figure 4.6, Figure 4.7 and Figure 4.8.

4.4 Sequencing of PCR Product

Once the PCR reaction is complete, the sequence of the antisense strand needs to be determined. This sequence will indicate the ratio of natural nucleotides incorporated across from the pyrazole analog. The sequence of the antisense strand will be determined using the Sanger Method as described in Section 1.4.3.

4.5 Results

4.5.1 PCR Reactions

The PCR reactions using the modified nucleoside analogs were successful as indicated by full-length product in the agarose gel. Figures 4.5-4.8 are analytical

agarose gels showing the results of each PCR reaction with the individual primers with each different enzyme.

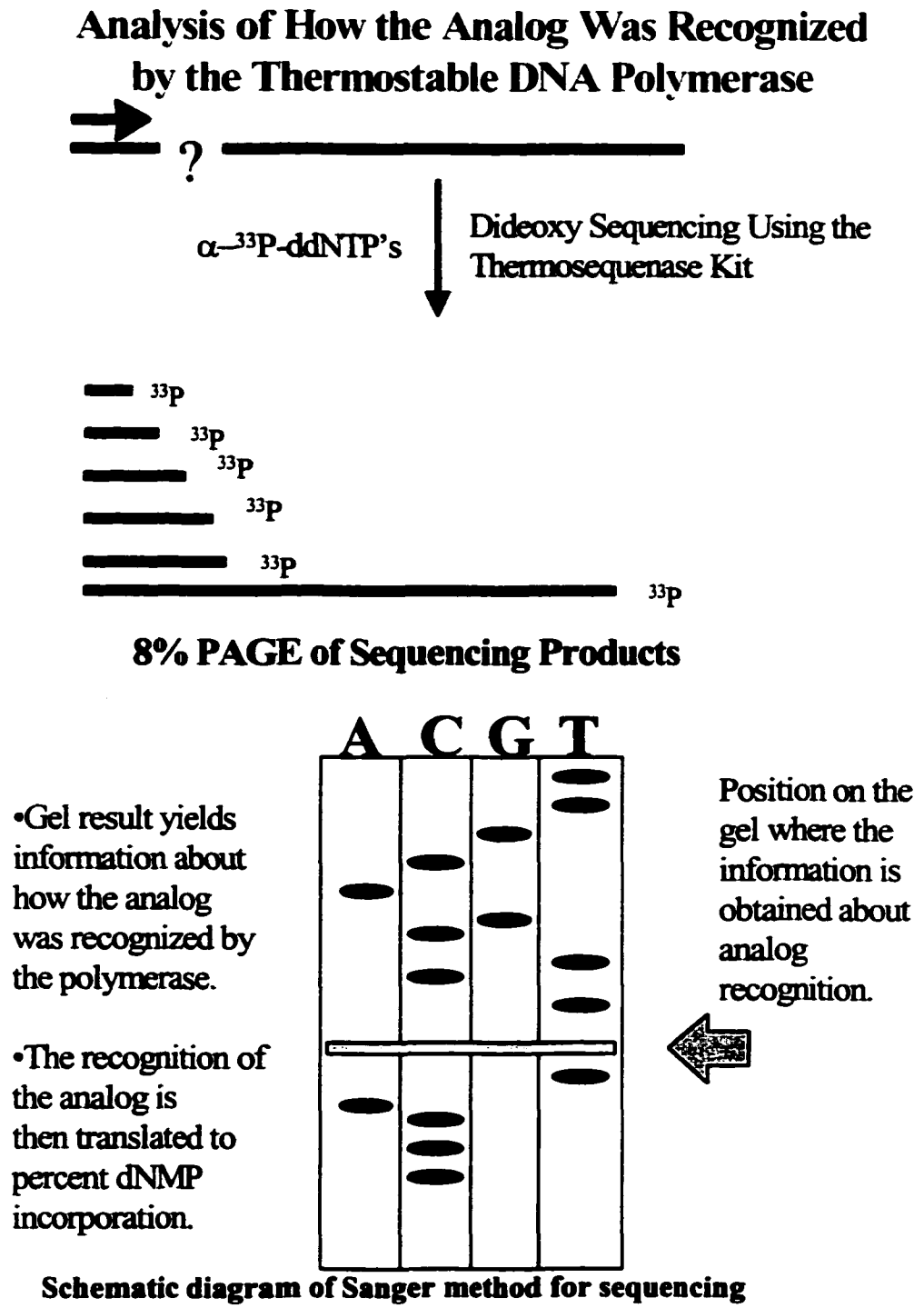


Figure 4.4 Schematic diagram of Sanger method for sequencing

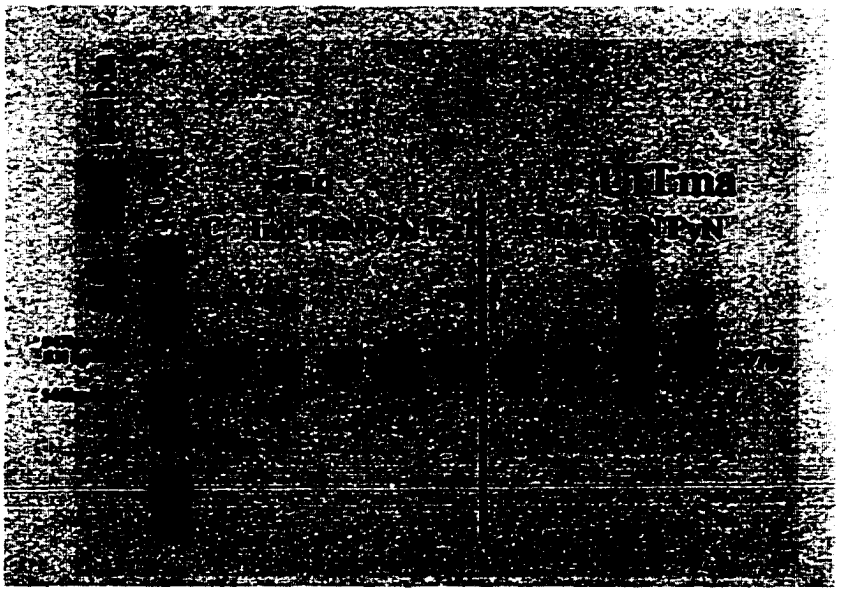


Figure 4.5 Agarose gel analyzing PCR reactions with AmpliTaq® and *UlTma* enzymes with control primer (C), 5-nitroindole (Ind) primer, 4-nitropyrazole (PzN) primer, 3-nitropyrrole (PyN) primer and 4-(2-thiazolyl)-pyrazole (PzT) primer

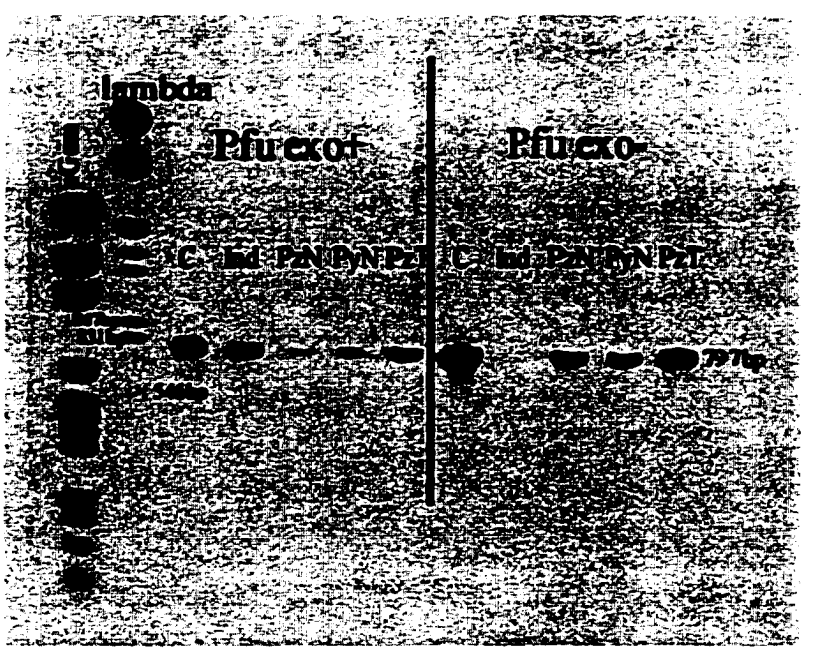


Figure 4.6 Agarose gel analyzing PCR reactions with *Pfu* exo+ and P*fu exo-* Enzymes with control primer (C), 5-nitroindole (Ind) primer, 4-nitropyrazole (PzN) primer, 3-nitropyrrole (PyN) primer and 4-(2-thiazolyl)-pyrazole (PzT) primer

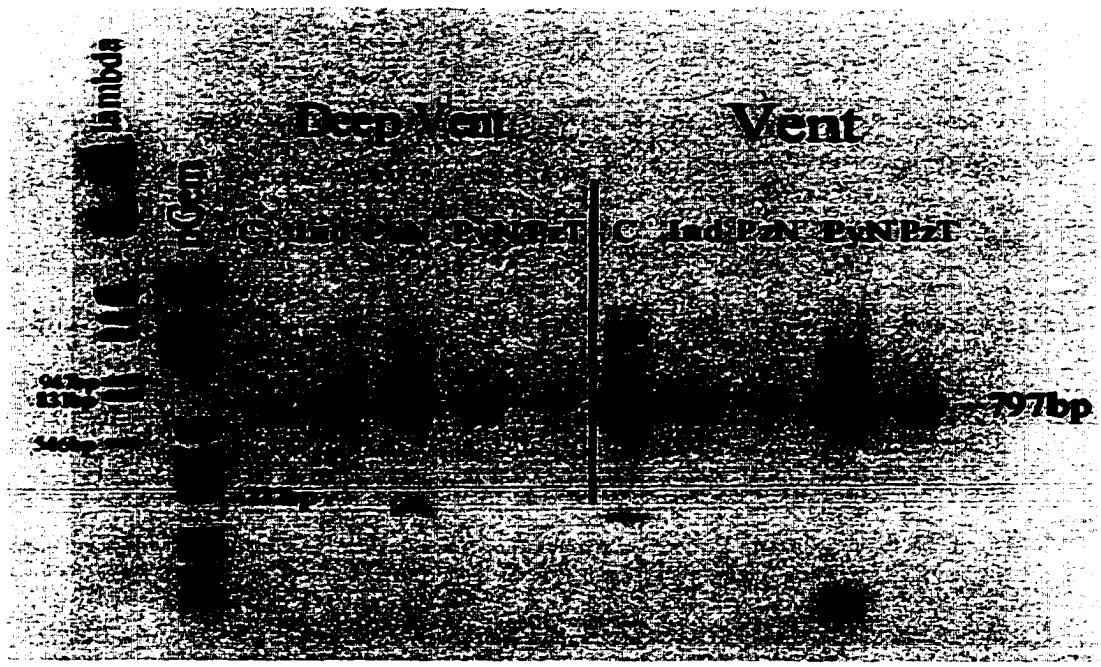


Figure 4.7 Agarose gel analyzing PCR reactions with Deep Vent and Vent enzymes with control primer (C), 5-nitroindole (Ind) primer, 4-nitropyrazole (PzN) primer, 3-nitropyrrole (PyN) primer and 4-(2-thiazolyl)-pyrazole (PzT) primer

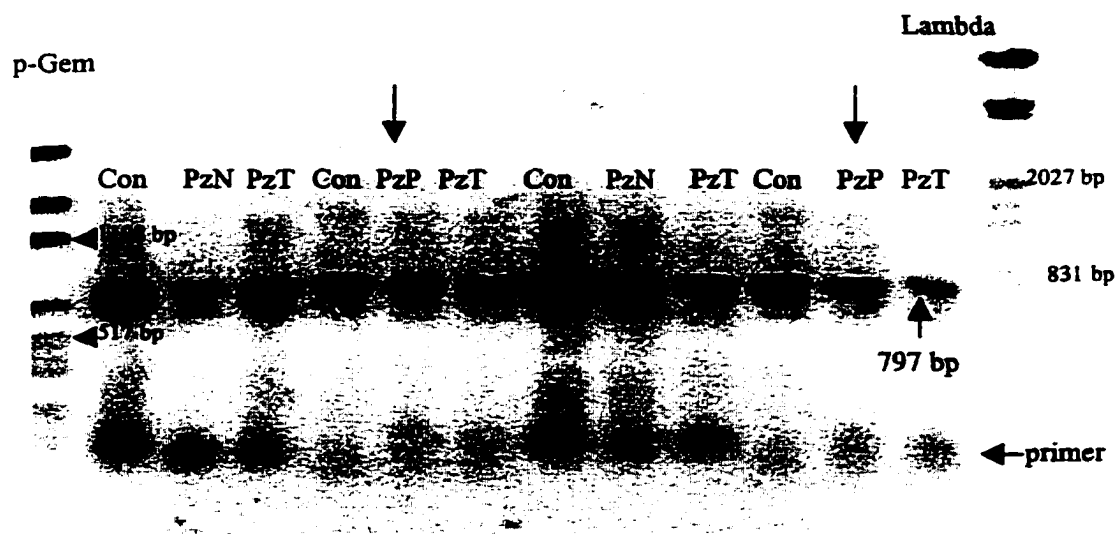


Figure 4.8 Agaorse gel analyzing PCR reactions with AmpliTaq® enzyme with control Primer (Con), 4-nitropyrazole primer (PzN), 4-propynyl pyrazole primer (PzP), and 4-(2-thiazolyl)pyrazole primer (PzT)

4.5.2 Sequencing of PCR Products

Sequencing of the PCR products was successful using Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit (Amersham/USB). Analysis of the sequencing reaction again used gel electrophoresis yet acrylamide/bisacrylamide was used as the separation medium. This is more appropriate for shorter strands of DNA when compared with agarose. When using the PCR product from the reaction with the control sense primer, the position of interest should show G. With modified primers other bases can be incorporated as shown in Figure 4.8. In most instances with non-proofreading enzymes, T was incorporated across from PzN and PzT. Proofreading enzymes tended to remove the analogs but not with total efficiency.

4.5.3 Quantitation of Acrylamide Gel Sequencing Results

The results of this research are ultimately determined by quantitation of the gels of the sequencing reactions. One of the actual gel images used for quantitation is included for each primer and each enzyme in Figure 4.8 *Pfu* exo-, Figure 4.9 Ampli*Taq*®, Figure 4.10 Vent, Figure 4.10 Deep Vent, and Figure 4.12 *Pfu* exo+, and Figure 4.14 *UlTma*. The analysis of the gels was performed using ImageQuant by Molecular Dynamics. Table 4.2 contains the relative amount of dNMP incorporated into the antisense strand across from the modification. (Note: the gel image contains the complement to the base found on the PCR product.)

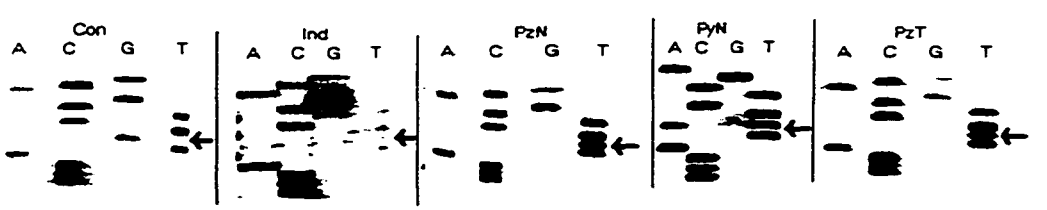


Figure 4.9 Sequencing of PCR products from *Pfu* exo- enzyme

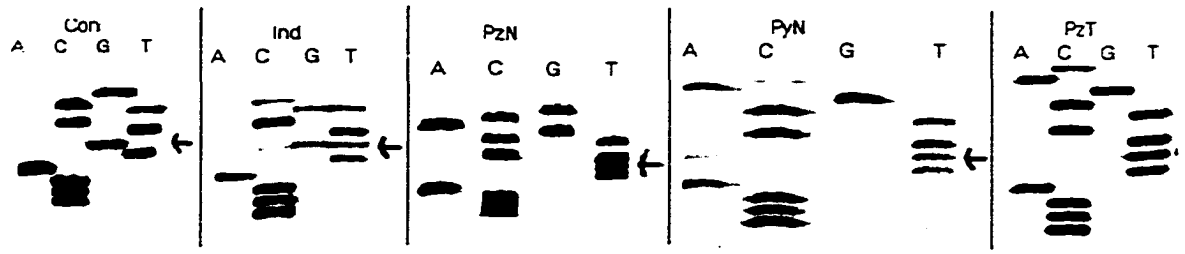


Figure 4.10 Sequencing gels of PCR products from Ampli*Taq*®

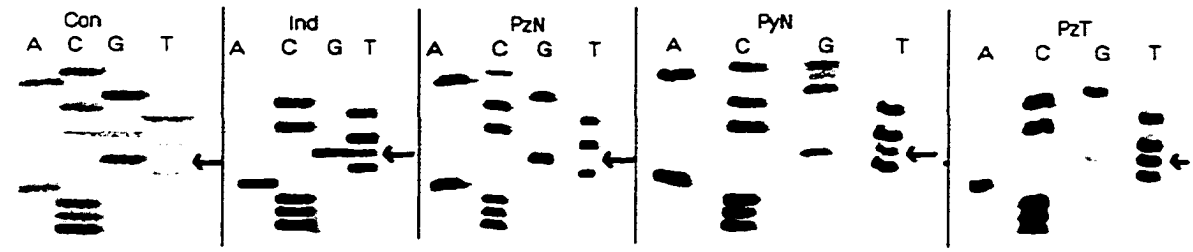


Figure 4.11 Sequencing gels of PCR products from Vent enzyme

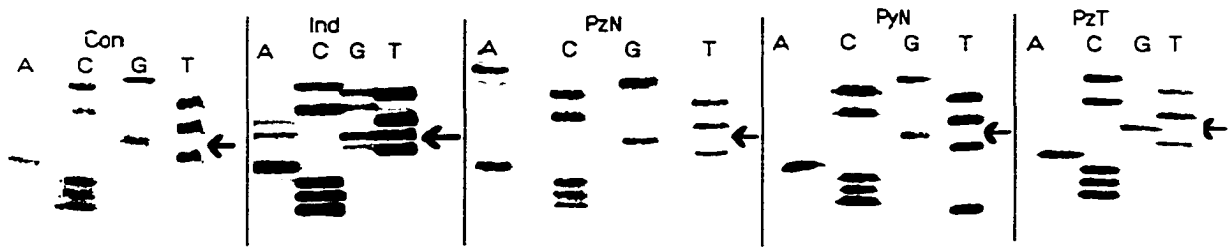


Figure 4.12 Sequencing of PCR products from Deep Vent enzyme

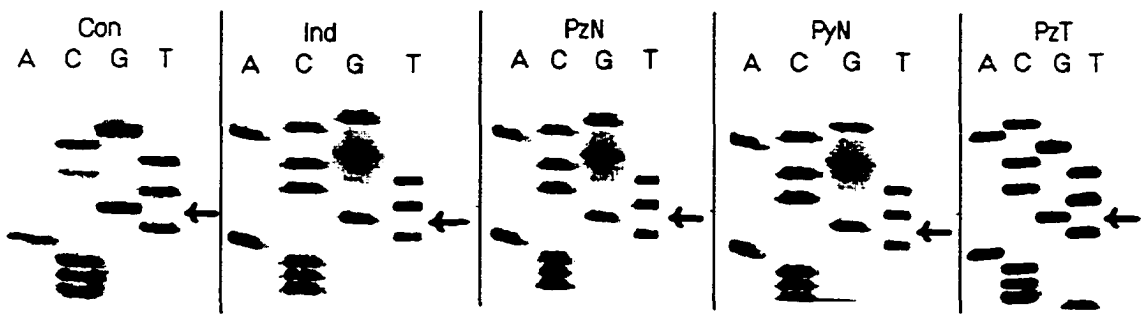


Figure 4.13 Sequencing of PCR products from *Pfu* exo+ enzyme

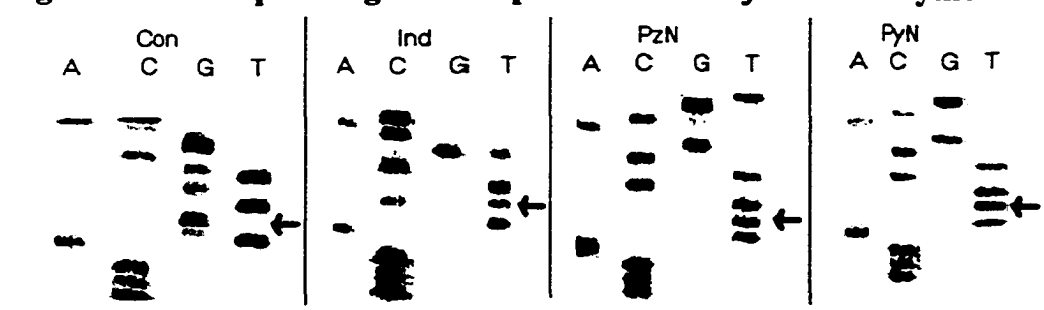


Figure 4.14 Sequencing Gel of PCR products from *UlTma* enzyme

Table 4.2 % dNMP Incorporated into PCR Reactions Using Primers Containing Modification in place of G

Enzyme	AmpliTaq® (-)	*Pfu exo-*	Deep Vent (+)	Vent(+)	*Pfu exo+*	*UlTma* (+)
Primer						
Control (G)	A 3 (1) C 96 (2) G 1 (1) T 0 (0)	A 7.5 (1) C 91.5 (2) G 0 (0) T 1 (1)	A 1 (2) C 99 (2) G 0 (0) T 0 (0)	A 0 (0) C 82 (8) G 18 (8) T 0 (0)	A 2 (2) C 98 (2) G 0 (0) T 0 (0)	A 0 (0) C 92 (10) G 1 (1) T 7 (11)
Ind	A 51 (4) C 22 (20) G 26 (20) T 1 (1)	A 15 (3) C 27 (1) G 29 (1) T 29 (5)	A 31 (37) C 65 (37) G 1 (2) T 3 (2)	A 48 (15) C 52 (15) G 0 (0) T 0 (0)	A 3 (4) C 96 (3) G 1 (2) T 0 (0)	Not determined
PzN	A 94 (2) C 4 (1) G 2 (1) T 0 (0)	A 99 (2) C 0 (0) G 1 (2) T 0 (0)	A 15 (13) C 85 (13) G 0 (0) T 0 (0)	A 7 (7) C 93 (7) G 0 (0) T 0 (0)	A 10 (6) C 77 (10) G 4 (4) T 9 (8)	A 99 (1) C 1 (1) G 0 (0) T 0 (0)
PyN	A 83 (4) C 2 (2) G 3 (2) T 12 (8)	A 45 (13) C 18 (8) G 10 (3) T 25 (14)	A 7 (12) C 87 13) G 1 (2) T 5 (4)	A 51 (22) C 43 (21) G 2 (2) T 5 (3)	A 2 (1) C 98 (2) G 0 (0) T 0 (0)	A 97 (6) C 0 (1) G 3 (5) T 0 (0)
PzT	A 93 (6) C 5 (5) G 1 (1) T 0 (0)	A 100 (0) C 0 (0) G 0 (0) T 0 (0)	A 28 (2) C 63 (2) G 6 (2) T 3 (2)	A 80 (6) C 18 (9) G 2 (4) T 0 (0)	A 8 (4) C 91 (5) G 1 (1) T 0 (0)	A 96 (8) C 4 (8) G 0 (0) T 0 (0)
25%A/ C/G/T	A 26 (8) C 32 (5) G 23 (8) T 19 (5)					

4.6 Discussion

Non-proofreading enzymes *AmpliTaq*® and *Pfu-* mainly paired an A with the pyrazole analogs. This would agree with the thermodynamic analysis indicating the pyrazole analogs show a preference for pairing with A. Bergstrom and coworkers report mainly A incorporation across from an abasic site.[1] Our research found mixtures of the nucleotides incorporated across from **Ind** and **PyN**. The ratio of nucleotides for **PyN** is very similar to those found by Bergstrom.[1] A possible explanation of a purine/pyrimdines mixture is the ability of the base to rotate about the glycosidic bond in an anti-syn configuration, allowing for size accommodations depending on the pairing partner. The pyrazole analogs are unable to have this rotation due to the placement of the substituents on the 4-position. The position of the substituents, although the entire molecule is no larger than indole, would disturb the complementary chain is the other conformation is adopted.

Two control experiments were performed. This first contained the natural complement, G, in the primer to the base in the template strand. The first control allows us to determine whether the polymerase and all the other reagent were correctl added. This also provides a contrasting base in the sequencing reaction as to what base we expected to be incorporated. If our analogs were expected to direct for a C in the PCR reaction it would be difficult to detect. With the proof reading enzymes, it the modification was removed from the initial primer and replaced with the natural complement to the template, a G is indicated in the sequencing reaction. If the modification is not removed then something other than G normally has been

observed. The second control experiment used each natural base in place of the modification in equal amounts provided close to 25% with error of each base. (Noted in Table 4.2 as 25%A/C/G/T) The C base had a slightly higher amount but this is the natural complement allowing possibly for better annealing as well as ease of extension off of a mismatch. Proofreading enzymes, Deep Vent, Vent and *Pfu+*, removed all of the analogs but with varying efficiency. If the modifications are not removed an A is incorporated across from the pyrazole analog. The exception to the proofreading enzymes is *UlTma*. *UlTma* does not remove modifications and incorporates an A with the analogs. These results indicate the pyrazole analogs direct for exclusively A incorporation with the enzymes studied whereas **Ind** and **PyN** do not exclusively direct for one of the natural bases.

Another explanation of incorporation of A across from the pyrazole analogs could be suggested as the "A-rule" possessed by some enzymes.[4] The "A-rule" suggests if an enzyme does not recognize the modification then an A will be incorporated as a "fail safe" and the polymerization will continue.[5] Polymerases extend off an A-mismatch with better efficiency than any of the other bases and thus "A" incorporation may reflect selective amplification of thie mismatch.[5,6] Exonuclease activity can remove abnormalities 5 bases back from the 3' end. This position of our modification in the primer was such that exonuclease activity could remove it. Depending on the enzyme, the non-natural analogs were tolerated to varying degrees. The sequencing does not indicate evidence of the polymerase skipping or looping out the pyrazole analog.

Insertion of A across from the pyrazole analogs seems to be favored. Although thermodynamic data supports A as the base favored across from the pyrazoles, the fail safe insertion of A by the polymerases cannot yet be ruled out. To determine if the pyrazole analogs are actually directing for A to be incorporated, kinetic experiments will need to be performed to determine the rate of insertion for the natural nucleotides against the pyrazole analogs.

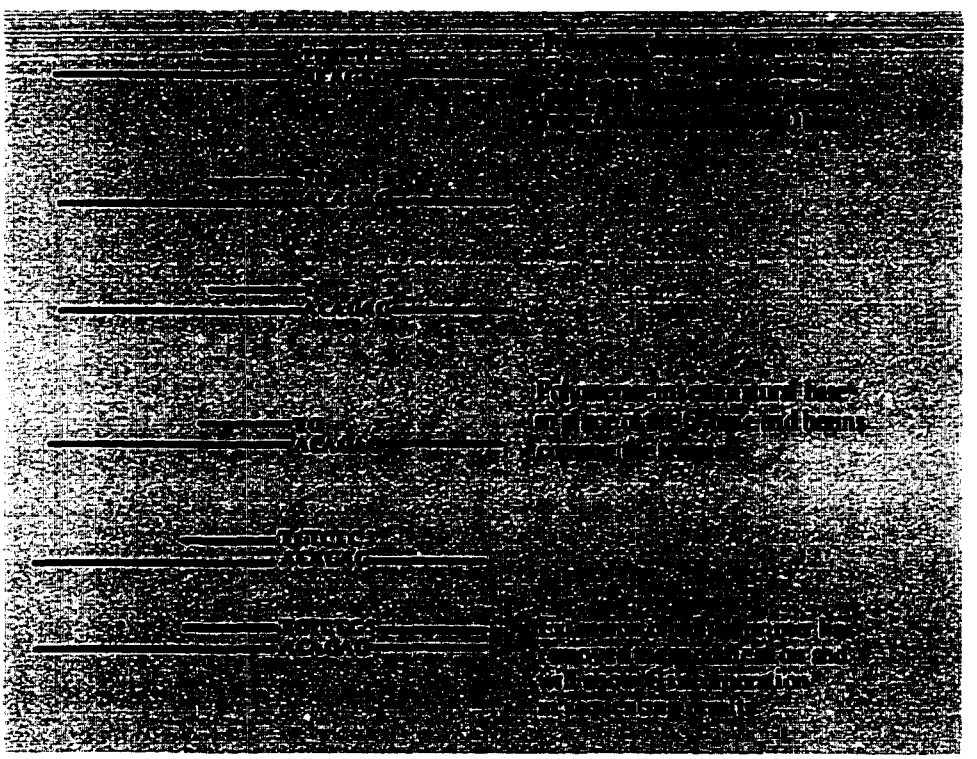


Figure 4.15 Schematic of 3'→5' Exonuclease Activity Removing the Modification Q from the Primer and Replacing with a Natural Nucleotide.

Related to the kinetic question is whether termination of the extension is occurring across from or immediately following the modification and to what extent. If termination were occurring, the position of termination would provide useful information about the "reading" efficiency by various polymerasesof the analogs. A shortcoming of the original experimental design does not allow for the determination

of whether the polymerase is inserting across from the analog and then stalling on the next insertion or if the polymerase is stalling at the modification. An assay which involves a fluorescently labeled antisense primer is being used to probe this topic. This primer results in shorter products (237 bp) which are simpler to analyze due to their size as well as the fluorescent label, which allows for more convenient analysis. According to preliminary results some termination is occurring.

4.7 Experimental

4.7.1 PCR Reactions

4.7.1.1 Preparation of Template

Pvu II restriction endonuclease was used to digest phisF-tac plasmid, generously provided by Davisson and coworkers,[1] at Purdue University into the hisF gene (797bp) (Figure 4.3) following the protocol provided by Promega with the enzyme (Table 4.3). The *Pvu* II recognition site is 5'-CAG^CTG-3'. Once the reagents were added the reaction was incubated at 37 °C for 1 hour and then diluted 10 fold to a final concentration of 20 fmol/ μL.

Table 4.3. Conditions for Digesting phis-F-tac plasmid with Pvu II

Reagent	Stock Concentration	Volume μL	Final Concentration (50μL)
H_2O		41.75	
phis-F tac	1.7μg/μL	1.25	42.5ng/μL
PVU II	10 Units/ μL	2	20 Units
10X NEBuffer 2	500mM NaCl 100mM Tris-HCl 100mM $MgCl_2$ 10mM DTT	5	50mM NaCl 10mM Tris-HCl 10mM $MgCl_2$ 1mM DTT

4.7.1.2 Synthesis of Primers

Most primers were synthesized by Midland Certified Reagent Company using standard automated DNA chemistry and purified by AE-HPLC. Several primers were synthesized at LSU. These were prepared using the same procedure as for the oligomers used in the thermal denaturation studies. See Section 2.3.2.

4.7.1.3 Sequence of Primers

Forward Sense Primer for PCR

5' AGG AAT TCA TAT GCT GGC AAA ACG CAT AAT CCC ATQ TCT C3'

Q= G control

4-nitropyrazole nucleotide,

4-propynylpyrazole nucleotide,

4-(2-thiazolyl)pyrazole nucleotide

5-nitroindole nucleotide

Antisense Primer for PCR

5' CGT ACC CGG GTT AAC ATA TCC TGA TCT CCA 3'

4.7.1.4 Conditions for PCR Reactions

The enzymes and respective buffers were purchased from commercial sources with one exception. *UlTma* was originally purchased from Perkin-Elmer and then a generous gift by Dr. Gefland supplied enough enzyme to complete the experiments. The commercial sources for each enzyme are: Ampli*Taq*® DNA polymerase (Perkin-Elmer), Vent DNA polymerase (New England Biolabs), Deep Vent DNA polymerase (New England Biolabs), Pfu DNA polymerase exo+, *Pfu* DNA polymerase exo-, and *UlTma* DNA polymerase (Perkin Elmer). The dNTP's were purchased from Roche. The following reagents were added into a 200 μL reaction tube. The reagents and

reaction tube were kept on ice until added to the thermal cycler heating block at 75 °C. The reaction was mixed thoroughly between each addition in the order listed. The cycle for the reaction included an initial warm up of 94 °C for 5 min., followed by 32 cycles of 95 °C for 30 seconds, 37 °C for 30 seconds and 70 °C for 1 min., and a final 10 min. extension at 70 °C and then temperature was lowered to 4 °C until further workup.

Table 4.4 Ampli*Taq*® PCR Conditions

Reagent	Stock Concentration	Volume µL	Final Concentration (100µL)
H_2O		52.5	
10X PCR Buffer II	500mM KCl, 100mM Tris-HCl, pH 8.3	10	50mM KCl, 10mM Tris-HCl
dNTP's	2mM total of each	10	200µM
Forward Primer (Modified)	10µM	10	1µM
Antisense Primer	10µM	10	1µM
$MgCl_2$ Solution	25mM $MgCl_2$	6	1.5mM $MgCl_2$
phis-F-tac digested plasmid	fmol quantities of template	1	fmol quantity
AmpliTaq® Polymerase	5 Units/µL	0.5	2.5 Units

Table 4.5 Vent PCR Conditions

Reagent	Stock Concentration	Volume µL	Final Concentration (100µL)
H_2O		58	
10X ThermoPol Reaction Buffer	100mM KCl, 100mM $(NH_4)_2SO_4$, 200mM Tris-HCl (pH 8.8 @25 °C) 20mM $MgSO_4$ 0.1%Triton X-100	10	10mM KCl, 10 mM $(NH_4)_2SO_4$, 20mM Tris-HCl 2mM $MgSO_4$

Table 4.5 continued

dNTP's	2mM total of each	10	200μM
Forward Primer (Modified)	10μM	10	1μM
Antisense Primer	10μM	10	1μM
phis-F-tac digested plasmid	fmol quantities of template	1	fmol quantity
Vent DNA Polymerase	2 Units/μL	1	1 Unit

Table 4.6 Deep Vent PCR Conditions

Reagent	Stock Concentration	Volume μL	Final Concentration (100μL)
H_2O		58	
10X ThermoPol Reaction Buffer	100mM KCl, 100mM $(NH_4)_2SO_4$, 200mM Tris-HCl, (pH 8.8 @25 °C) 20mM $MgSO_4$ 0.1%Triton X-100,	10	10mM KCl, 10 mM $(NH_4)_2SO_4$, 20mM Tris-HCl, 2mM $MgSO_4$,
dNTP's	2mM total of each	10	200μM
Forward Primer (Modified)	10μM	10	1μM
Antisense Primer	10μM	10	1μM
phis-F-tac digested plasmid	fmol quantities of template	1	fmol quantity
Deep Vent DNA Polymerase	2 Units/μL	1	1 Unit

Table 4.7 Cloned *Pfu* exo+ PCR Conditions

Reagent	Stock Concentration	Volume μL	Final Concentration (100μL)
H_2O		58	
dNTP's	2mM total of each	10	200μM
Forward Primer (Modified)	10μM	10	1μM

Table 4.7 continued

10X ThermoPol Reaction Buffer	100mM KCl, 100mM $(NH_4)_2SO_4$, 200mM Tris-HCl (pH 8.8 @25 °C) 20mM $MgSO_4$ 1%Triton X-100, 1mg/ml nuclease-free bovine serum albumin	10	10mM KCl, 10 mM $(NH_4)_2SO_4$, 20mM Tris-HCl 2mM $MgSO_4$
Antisense Primer	10μM	10	1μM
phis-F-tac digested plasmid	fmol quantities of template	1	fmol quantity
Cloned *Pfu*+ DNA Polymerase	2.5 Units/μL	1.25	1 Unit

Table 4.8 Recombinant *Pfu* exo- PCR Conditions

Reagent	Stock Concentration	Volume μL	Final Concentration (100μL)
H_2O		58	
10X ThermoPol Reaction Buffer	100mM KCl, 100mM $(NH_4)_2SO_4$, 200mM Tris-HCl (pH 8.8 @25 °C) 20mM $MgSO_4$ 1%Triton X-100, 1mg/ml nuclease-free bovine serum albumin	10	10mM KCl, 10 mM $(NH_4)_2SO_4$, 20mM Tris-HCl 2mM $MgSO_4$
dNTP's	2mM total of each	10	200μM
Forward Primer (Modified)	10μM	10	1μM
Antisense Primer	10μM	10	1μM
phis-F-tac digested plasmid	fmol quantities of template	1	fmol quantity
Recombinant *Pfu*- DNA Polymerase	2.5 Units/μL	1.25	1 Unit

Table 4.9 ***UlTma* PCR Conditions**

Reagent	Stock Concentration	Volume µL	Final Concentration (100µL)
H_2O		52.5	
10X PCR Buffer	100mM KCl, 100mM Tris-HCl, pH 8.8)	10	10mM KCl 10mM Tris-HCl
dNTP's	2mM total of each	10	200µM
Forward Primer (Modified)	10µM	10	1µM
Antisense Primer	10µM	10	1µM
$MgCl_2$	25 mM $MgCl_2$	6	1.5 µM $MgCl_2$
phis-F-tac digested plasmid	fmol quantities of template	1	fmol quantity
UlTma	6 Units/µL	0.5	3 Unit

4.7.1.5 Gel Electrophoresis of PCR Products.

The products from the PCR reaction were separated and visualized using agarose gel electrophoresis (BioRad). The samples were loaded onto a 1% (mol / v) agarose gel with a buffer of Tris-Acetate-EDTA (TAE)(0.04 M Tris-Acetate, 0.001 M EDTA). The gel was prepared by dissolving 1 g of agarose in 75 mL of 1X TAE (0.04 M Tris-Acetate, 0.001M EDTA) by heating. Once all the agarose was dissolved, 3 µL of ethidium bromide was added for later visualization, the warm solution was poured into the electrophoresis tray, allowed to cool and solidify into a whitish gel. Next, the gle was placed into the electrophoresis tray with buffer and loaded with equal volumes of the samples and a 2X Load buffer (0.08% bromophenol blue, 0.08% xylene cyanol FF and 5% Ficoll (Type 400)). The gel was run with 100V applied for 30 minutes or until the first dye marker traveled half the distance of the gel. The ethidium bromide intercalated into the DNA and allowed for

visualization using UV light. Two DNA size markers from Promega were used to detemine size of PCR products, pGem ®(Catalog # g1741) and Lambda DNA/*Eco*R I + *Hind* III Markers (Catalog # 1731), abbreviated as lambda in the figures. The source for pGem® DNA markers are made by digesting pGem®-3 DNA to completion using *Hinf* I, *Rsa* I and *Sin* I. The marker called lambda results from the complete digestion of Lambda DNA with *Eco*R I and *Hind* III enzymes.

4.7.1.6 Purification of PCR Products

Wizard PCR preps (Promega) were used for purification of the PCR products following the manufacture's protocol.

4.7.2 Sequencing Reactions

4.7.2.1 Synthesis of DNA Primer for Sequencing Reactions

The primer for sequencing was synthesized by Midland Certified Reagent Company and purified by AE-HPLC.

Sequence 5' GAA TTC ATA TGC TGG C 3'

4.7.2.2 Reaction Conditions for Sequencing Reactions

The sequencing primer (pmol) along with the purified PCR products (fmol) were added to the reagents provided by Thermosequenase™ kit (Amersham). The manufacturer's protocols were followed for sequencing. The thermal cycle required for the reaction was provided by the kit.

4.7.2.3 Gel Electrophoresis of Sequencing Reactions

The quenched reaction products were analyzed under denaturing conditions (7 M urea 50-55 °C) using gel electrophoresis (8%, 19:1 acrlyamide:bisacrylamide)

under constant voltage (1100 V) on a BioRad Sequi-Gen GT System (0.4 mm x 38 cm x 30 cm). The gels were cast using the precision cast assembly provided by Bio-Rad and equilibrated for 1h before loading. The 4 (A, C, G, T) samples were heated to 70 °C and then 4.0 μL was immediately added to the gel. A positive displacement pipette was used to assure the exact same amount of sample was added to each of the four lanes. If different amounts were added the quantitation would not be acceptable. The gels were run until the bromophenol blue in the loading buffer (provided by kit) had run 75% the length of the gel. The gels were then sandwiched between blotting paper and Saran Wrap® and dried for 2 h at 80 °C under vacuum on a Bio-Rad 583 gel dryer. Once the gel had dried and cooled, it was placed on the imaging plate (Molecular Dynamics) for 24 h and scanned using Image-Quant (Molecular Dynamics) software

4.7.2.4 Quantitation of Images from Sequencing Gels

Anaylsis of the image was performed using ImageQuant. All four lanes at the position the modification was analyzed. To quantitate, a rectangle was drawn around each lane at the position of interest. A second rectangle was placed in a position of representative. Each lane was corrected for the background. The software then analyzed the four lanes for relative intensities in the lanes. This analysis was performed on alteast three separate sequencing reactions. The average of the anaylsis is reported with the standard deviation.

4.8 References

1) Hoops, G. C.; Zhang, P.; Johnson, W. T.; Paul, N.; Bergstrom, D. E.; Davisson, V. J. *Nucleic Acids Res.* **1997**, *25*, 4866-4871.

2) Klem, T. J.; Davisson, V. J. *Biochemistry* **1993**, *32*, 5177-5186.

3) Cline, J.; Braman, J. C.; Hogrefe, H. H. *Nucleic Acids Res.* **1996**, *24*, 3546-3551.

4) Strauss, B. S. *BioEssays* **1991**, *13*, 79-84.

5) Goodman, M. F.; Creighton, S.; Bloom, L. B.; Petruska, J. *Critical Reviews in Biochemsitry and Molecular Biology* **1993**, *28*, 83-126.

6) Cai, H.; Bloom, L. B.; Erijia, R.; Goodman, M. F. *J. Biol. Chem.* **1993**, *268*, 23567-23572.

Chapter 5
Synthesis of 2-thiazole C-nucleoside

5.1 Design

Examination of the thermodynamic results for 4-(2-thiazolyl)-pyrazole (**PzT**) (Chapter 3) led to an interest in synthesizing 2-thiazole C-nucleoside. The thermodynamic data showed **PzT** to be the best of the four pyrazole analogs studied. The thiazole substituent provided increased T_m's with the self-complementary sequences when compared to the other three substituents. We would like to investigate thiazole as a main heteroaromatic unit in thermodynamic studies as well as enzymic studies. Substituted thiazole nucleosides have been studied before mainly in RNA for their antiviral activity.[1,2] Unsubstituted thiazole nucleoside had not been previously been synthesized in the literature.

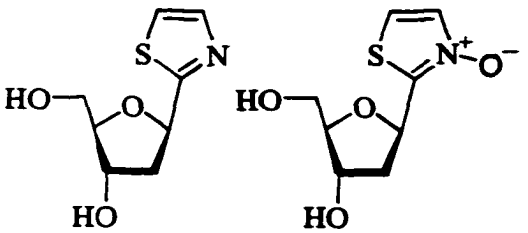


Figure 5.1 Structure of thiazole nucleoside and thiazole N-oxide nucleoside

Two design features with thiazole as the main heteroaromatic unit are a C-nucleoside and the introduction of both positive and negative charge in the ring. (Figure 2.3) A third design aspect with thiazole offers the ability to prepare an N-oxide. It has been shown that modification of the minor groove binding site of the Klenow fragment markedly decreased the DNA-binding affinity and the rate of DNA synthesis.[3-5] It has been proposed by Kool and others that the C-2-oxygen on

pyrimidines or N-3 in purines in the minor groove is a key feature for recognition by DNA polymerases.[6] For example, the nucleoside triphosphate of 3-nitropyrrole which lacks a minor groove hydrogen-bond donor, was shown to either inhibit or terminate polymerase extension with various polymerases.[7] The thiazole N-oxide would offer a non-hydrogen-bonding analog that allows for the placement of an oxygen in the minor groove. Comparison of enzyme behavior with the N-oxide versus the thiazole compound could provide insight into this theory.

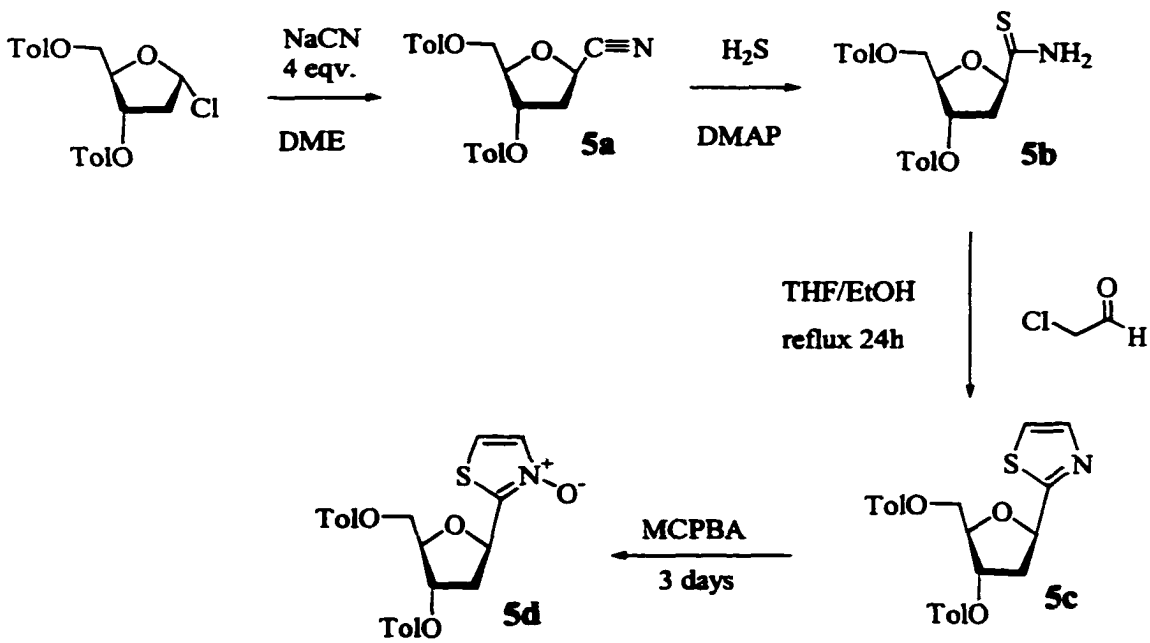
The synthesis of the thiazole compound (**5c**) and N-oxide compound (**5d**) are complete. The crystal structure of deprotected **5c** has been obtained. Incorporation into oligonucleotides has not been attempted. The stability of the oxide to standard automated DNA chemistry needs to be determined.

5.2 Synthesis

The synthetic approach developed for thiazole C-nucleoside is similar to the synthesis of compounds **1a** and **1b** in Chapter 2. The initial difficulty is the synthesis of a C-nucleoside. C-nucleosides are much more difficult to when compared to N-nucleosides. Sodium cyanide is dissolved in **freshly** distilled dimethoxyethane (DME) and of 3,5-bis-*O*-*p*-toluoyl-β-D-2-deoxyribose chloride was added portion wise and allowed to stir under inert atmosphere overnight forming yellowish salts on the bottom of the reaction flask. This solvent system according to the literature afforded a 2:1 ratio of β:α.[8] Using freshly distilled DME we found almost 95% of the β anomer. The cyano sugar required only an aqueous extraction to provide material pure enough for the next reaction. The next series of

synthetic steps required the conversion of the cyano group into thioamide using H_2S (g) followed by cyclization using chloroacetaldehyde. This series of steps is a modification of the synthesis of 4-thiazolecarboxamide. [9] This conversion of a cyano group into thioamide is a very common synthetic procedure condensing H_2S (g) with the cyano functional group. The crude product from this conversion required purification by column chromatography. The final cyclization step required refluxing chloroacetaldehye with the thioamide in a mixture of tetrahydrofuran/ethanol.[10] Attempts to cyclize without heat were not successful. The N-oxide of **5c** was prepared with exclusive oxidation on the ring nitrogen without oxidation of the ring sulfur as proposed by Hansen and coworkers using *m*-chloroperoxybenzoic acid (MCPBA).[11] This reaction required several days with only a modest yield. (Scheme 5.1)

Scheme 5.1 Synthesis of 2-thiazole Protected C-nucleoside and Thiazole-N-oxide C-nucleoside



5.3 Experimental

5.3.1 5-anhydro-3-deoxy-4,6-di-*O*-*p*-toluoyl-β-D-*ribo*-hexononitrile (5a)

Sodium cyanide (1.49 g, 0.03 mol) was dissolved in DME (100 mL) freshly distilled from Na (s) and benzophenone under Ar. This solution stirred for 30 minutes followed by the addition of 3,5-bis-*O*-*p*-toluoyl-β-D-2-deoxyribose chloride (2.97 g, 7.6 mmol) (Scheme 2.1). The cloudy white solution turned yellowish with a small amount of solid precipitate when left overnight to stir under inert atmosphere. The DME was removed under reduced pressure. Next, the crude product was redissolved in ethyl acetate and washed with saturated sodium bicarbonate, water and brine, dried and filtered. The solvent was again removed under reduced pressure and the yellowish oil (2.56 g, 67%) was used directly in the next reaction. Characterization by ^{1}H NMR (250 MHz, $CDCl_3$), δ 7.97 (d, J=8, 2H aromatic), 7.85 (d, J=8, 2H aromatic), 7.20 (d, J=8, 4H, aromatic), 5.60 (m, 1H, H3), 4.90 (t, J=7.7, 1H, H1α), 4.52 (m, 3H, H4, H5, H5'), 2.82 (m, 1H, H2) 2.52 (m, 1H, H2'), 2.39 (s, 6H, Ar-CH_3). The ratio of β:α is 95:5. Characterization of α anomer by ^{1}H NMR (250 MHz, $CDCl_3$), δ 7.97 (d, J=8, 2H aromatic), 7.85 (d, J=8, 2H aromatic), 7.20 (d, J=8, 4H, aromatic), 5.60 (m, 1H, H3), 5.04 (d, J=7Hz, 1H, H1β), 4.52 (m, 3H, H4, H5, H5'), 2.82 (m, 1H, H2) 2.52 (m, 1H, H2'), 2.39 (s, 6H, Ar-CH_3).

5.3.2 5-anhydro-3-deoxy-4,6-di-*O*-*p*-toluoyl-β-D-*ribo*-hexonothiamide (5b)

5-anhydro-3-deoxy-4,6-di-*O*-*p*-toluoyl-β-D-*ribo*-hexononitrile, (**5a**), (2.33 g, 0.06 mmol) was dried in the reaction vessel overnight under vaccum. Next, DMAP (0.05 g, 0.41 mmol) was added, the pressure vessel cooled in a dry ice/acetone bath

and H_2S condensed into the vessel. Once a sufficient amount of gas had condensed, the vessel was seal and left to warm to room temperature overnight. The vessel was cooled again in a dry ice/acetone bath, opened and allowed to warm to room temperature. The orangish oil was dissolved in DCM and transferred to a second flask. The DCM was removed under reduced pressure. The crude reaction was then dissolved in ether and washed with sodium bicarbonate, water and brine, dried with sodium sulfate, filtered, and the solvent removed under reduced pressure. The β anomer was isolated using silica with 95:5 $CHCl_3$/EtOAc ($R_f = 0.03$) in 27% yield (0.68 g, 1.6 mmol). ^{1}H NMR (250 MHz, $CDCl_3$), δ 8.35 (s, 1H, amide), 7.85 (m, 4H aromatic), 7.57 (s, 1H, amide), 7.20 (d, J=8, 4H, aromatic), 5.60 (m, 1H, H3), 4.90 (t, J=7.7, 1H, H1), 4.52 (m, 3H, H4, H5, H5'), 2.82 (m, 1H, H2) 2.52 (m, 1H, H2'), 2.39 (s, 6H, Ar-CH_3)

5.3.3 2-(3',5'-Bis-*O*-*p*-toluoyl-β-D-2'-deoxyribosyl)-thiazole (5c)

5-anhydro-3-deoxy-4,6-di-*O*-*p*-toluoyl-β-D-ribo-hexonothiamide, (**5b**), (1.16 g, 3.0 mmol) was dissolved in THF (33 mL), ethanol (16 mL) followed by chloroacetaldehyde (50% wt. in H_2O) (3.18 mL, 25 mmol). This solution was left at reflux for 24 h followed by addition of a second portion of chloroacetaldehyde (3.18 mL, 25 mmol) with refluxing for 24 h. The solvents were removed under reduced pressure. The residue was dissolved in ethyl acetate washed with saturated sodium bicarbonate, water and brine, dried with sodium sulfate, filtered and the solvents were removed under reduced pressure. Isolation of the β anomer in 42% yield (0.55 g, 1.2 mmol) required silica with hexanes/ethyl acetate. ^{1}H NMR (250 MHz, $CDCl_3$, 7.85 (m, 4H aromatic, 1H H4), 7.75 (s, 1H, H5), 7.20 (d, J=8, 4H, aromatic), 5.60

(m, 2H, H3', H1'), 4.52 (m, 3H, H4, H5', H5"), 2.82 (m, 1H, H2") 2.52 (m, 1H, H2'), 2.39 (s, 6H, Ar-CH_3)

5.3.4 2-(3',5'-Bis-*O*-*p*-toluoyl-β-D-2'-deoxyribosyl)-thiazole-3-oxide (5d)

2-(3',5'-Bis-*O*-*p*-toluoyl-β-D-2'-deoxyribosyl)-thiazole (**5c**) (160 mg, 0.39 mmol) was dissolved in 2 mL of ethyl acetate followed by addition of *m*-chloroperoxybenzoic acid (MCPBA) (152 mg, 0.78 mmol). The reaction was allowed to stir for 2 days, followed by addition of a second portion of MCPBA (114 mg, 0.58 mmol) and left to stir for another 24 h. The reaction was quenched by ethyl acetate (25 mL) and sodium bicarbonate (25 mg). The mixture was filtered through a pad of solid sodium bicarbonate. The bicarbonate pad was rinsed with ethyl acetate followed by 20% methanol/ethyl acetate until no UV active compounds were eluting as indicated by TLC. The solvents were removed under reduced pressure and the residue was purified using silica with a gradient of 0-25% ethyl acetate in methanol. The clear oil was isolated in 42% yield. ^{1}H NMR (250 MHz, $CDCl_3$, 7.85 (m, 4H aromatic, 1H H4), 7.75 (s, 1H, H5), 7.20 (d, J=8, 4H, aromatic), 5.60 (m, 2H, H3', H1'), 4.52 (m, 3H, H4, H5', H5"), 2.82 (m, 1H, H2") 2.52 (m, 1H, H2'), 2.39 (s, 6H, Ar-CH_3)

5.4 References

1) Phelan, M. J.; Gabrielsen, B.; Kirsi, J. J.; Shannon, W. M.; Usser, M. A.; Barthel-Rosa, L.; Schubert, E. M.; Kini, G. D.; Robins, R. K. *Nucleosides Nucleotides* **1995**, *14*, 1315-1327.

2) Upadha, K.; DaRae, J.; Schubert, E. *Nucleosides Nucleotides* **1990**, *9*, 649-662.

3) Loeb, L. A.; Kunkel, T. A. *Annu. Rev. Biochem.* **1982**, *52*, 429.

4) Kuchta, R. D.; Mizrahi, V.; Benkovic, P. A.; Johnson, K. A.; Benkovic, S. J. *Biochemistry* **1987**, *26*, 8410-8417.

5) Kuchta, R. D.; Benkovic, P. A.; Benkovic, S. J. *Biochemistry* **1988**, *27*, 6716-6725.

6) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **1999**, *121*, 2323-2324.

7) Jacutin, S.; Zhang, A. J.; Russell, D. H.; Gibbs, R. A.; Burgess, K. *Nucleic Acids Res* **1997**, *25*, 5072-5076.

8) Kolb, A.; Huynh Dinh, T.; Igolen, J. *Bull. Soc. Che. Fr.* **1973**, *12*, 3447-3448.

9) Srivastava, P. C.; Robbins, R. K. *Heterocyclic Chem.* **1981**, *18*, 1659--1663.

10) Brandsma, L.; deJong, R. L.; ver Kruisse, H. D. *Synthesis* **1985**, 948.

11) Begtrup, M.; Hansen, L. L. *Acta. Chem. Scand.* **1992**, *46*, 372-383.

Chapter 6
Substitution of 5-Nitroindole Nucleotides During Oligonucleotide Synthesis

6.1 Introduction

Due to the repetitive cycling of automated DNA synthesis, it has been suggested by Dr. Donald Bergstrom at Purdue University as well as Dr. Lyn Meyers at Midland Certified Reagent Company, iodination of 5-nitroindole occurs by replacement of the nitro group in the 5 position or addition of iodine to the 3-position of the bicyclic ring. If substitution or addition occurred this might affect the enzyme behavior and recognition of primers containing this modification. Electrophilic aromatic substitution is the more common route for addition of halogens. Substitution or halogenation of aromatic systems can follow a radical cation mechanism when the appropriate activating substituents are present such as a nitro group.[1-3] The oxidation solution containing iodide, water and pyridine provides a source of $I^{\bullet}$ that could allow for radical collapse and therefore iodination of the indole ring. The presence of a radical cation might occur if the 5-nitroindole ring were to undergo oxidation (oxidation potential 1.59 eV[4]) due to excess reagents not totally removed from the column. The combination of the aromatic substitution and radical cation and addition could alter the characteristics of the indole ring when exposed to enzymes.

6.2 Discussion

Due to the position of 5-nitroindole in the primer sequence used in Section 4.7.1.3, the nucleoside is exposed to "excess" amount of oxidation cycles, which may increase the chance for addition or substitution. The common oxidation

procedures utilize iodide solutions but a few alternative solutions are available. One oxidation alternative is t-butyl peroxide but as this produces radicals much easier than I_2 solution another alternative was needed. The reagent ultimately used for oxidation is (1S)-(+)-(10-camphorsulfonyl)-oxaziridine (CSO). Glen Research optimized the conditions for oxidation using CSO as compared to I_2 and found a concentration of 0.5 M in acetonitrile with a 3 minute oxidation time is as effective as a 0.02 M I_2 solution with a 30 second oxidation time.[5] Therefore this oxidation procedure was used when preparing the 5-nitroindole primer for reactions. It was decided the sequences synthesized for thermodynamic studies did not have the excessive exposure as the primer sequence and resynthesis of the sequences was not necessary.

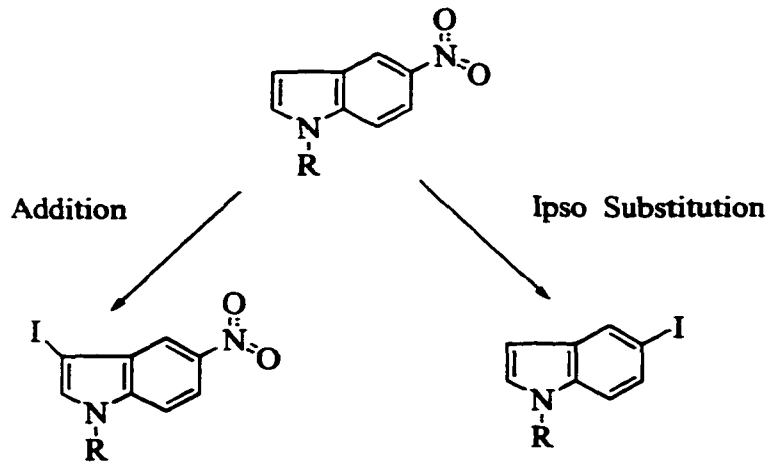


Figure 6.1 Iodination of 5-nitroindole

A series of test sequences were prepared in order to determine the extent of substitution. Two sequences composed of 4 natural bases with 5-nitroindole on the 5'end were prepared. One sequences was cleaved at the completion of the synthesis while a second synthesis was exposed to 35 more synthesis cycles yet no nucleotides were added. The two sequences were analyzed using MALDI-MS in

THA matrix by Tracey Simmons. The analysis of these spectra did not provide conclusive evidence. Most often a mass peak of +100 units was found in the test sequence. This would indicate the replacement the nitro group with iodine as a sodium salt. These results need to be repeated using a more complete set of internal standards for calibration. Further testing of 5-nitroindole nucleoside with exposure to 0.02 M I_2 solution over several days again did not provide sufficient evidence to support either addition or substitution.

NMR and mass spectrometry of the resulting solution did not show any evidence of replacement or addition. Investigation into the automated capping cycle and the possibility of any residual capping agents affecting the oxidation step might offer an alternative explanation.

Table 6.1 MALDI-MS results for iodination test sequences

Sequence	Oxidation Condition	$M+H^+$ calculated	$M+H^+$ found
5' CGTAQ 3'	0.02M I_2 solution	1515	1478 1486 1503
5' CGTAQ 3'	0.5M CSO solution	1515	1487 1504
5' CGTAQ 3' 35 more cycles	0.02M I_2 solution	1515	1488 1503 1583
5' CGTAQ 3' 35 more cycles	0.5M CSO solution	1515	1478 1487 1504

6.3 References

1) Hubig, S. M.; Jung, W.; Kochi, J. K. *J. Org. Chem.* **1994**, *59*, 6233-6244.

2) Eberson, L.; Hartshorn, M. P.; Radner, F.; Persson, O. *J. Chem. Soc., Perkin Trans. 2.* **1998**, 59-70.

3) Perrin, C. L. *J. Am. Chem. Soc.* **1977**, *99*, 5516-5518.

4) McCarley, R. L., Oxidation Potential for 5-nitroindole.

5) *Non-Aqueous Oxidation with 10-Camphorsufonyl-oxazirdine*; Glen Research Archive, 1999.

Vita

The author, Andrea Sapp Saurage, was born September 23, 1971 in Asheboro, North Carolina. She was raised in Burlington, North Carolina, and attended school in Burlington until high school where she attended and graduated from Foxcroft School in Middleburg, Virginia. She earned a bachelor of science degree in chemistry in 1993 from North Carolina State University, Raleigh, North Carolina, and followed with a bachelor of science degree in textile chemistry in 1994 from North Carolina State University, Raleigh, North Carolina. She began graduate school at Louisiana State University and Agricultural and Mechanical College in 1995 and joined the research group of Dr. Robert P. Hammer. She completed the degree requirements to receive the degree of Doctor of Philosophy in 1999. She is currently residing in Cary, North Carolina, with her husband, Kevin, and son, Evan.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Andrea Sapp Saurage

Major Field: Chemistry

Title of Dissertation: Synthesis, Thermodynamic Stability and Enzymic Behavior of Oligonucleotides Containing Pyrazole Nucleobase Analogs

Approved:

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

Date of Examination:

July 26, 1999